ADAPTIVE MODELING AND FORECASTING FOR HIGH-DIMENSIONAL TIME SERIES

LI BO

(B.Sc.(Hons) National University of Singapore)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Li Bo

12 August 2014

Thesis Supervisor

Ying CHEN Associate Professor; Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore.

Papers and Manuscript

Chen, Y. and Li, B. (2011). Forecasting Yield Curves in an Adaptive Framework, *Central European Journal of Economic Modeling and Econometrics*, 3(4): 237–259.

Chen, Y., Li, B. and Niu, L. (2013). A Local Vector Autoregressive Framework and its Applications to Multivariate Time Series Monitoring and Forecasting, *Statistics and Its Interface*, 6(4):499–509.

Chen, Y. and Li, B. (2014). Adaptive Functional Autoregressive Modeling for Stationary and Non-Stationary Functional Data, *Submitted and under revision.*

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my supervisor Professor Ying Chen for her patience, guidance, encouragement and most importantly, her enlightening ideas and valuable advice. I would like to thank Prof. Chen, not only for the knowledge passed on but also for the passion she has demonstrated in doing research, which have been tremendously helpful to me throughout these years. The gratitude I owe not only arises from the formal academic supervision that I receive; at the same time, it has also been due to Prof. Chen's continuous support for all aspects of my PhD study, in particular on possible research related opportunities granted to me. Here, it is my honor to take this opportunity to extend my hearty gratitude to my dear supervisor for all the memorable moments, both exciting and challenging sometimes, that she has shared with me.

I would also like to thank Professor Wolfgang Härdle for generously sharing his

ideas and his invited visit to Center for Applied Statistics and Economics in Berlin where the chances of exchanging research ideas and broadening my knowledge scale have been granted to me. The alerting and enlightening talks and discussions with Prof. Härdle have been rewarding and very helpful. Besides, my friends whom I made the acquaintance of during the visit to Berlin have made the visit very interesting and joyful. I would like to thank Weining Wang and Lining Yu for their thoughtful reception and those interesting discussions we have shared together.

Meanwhile, it is my pleasure to thank Professor Yingcun Xia and Professor Wei Liem Loh for many helpful conversations on both academic and non-academic affairs. I would also like to extend my gratitude to Prof. Xia, Professor Jialiang Li and Professor Kian Guan Lim from SMU for being my PhD thesis examiners.

Besides, I owe many thanks to my peer PhD students who have devoted their time and attention for the discussions we have had together. Their suggestions are of great help, which facilitates the accomplishment of the projects discussed in this thesis. At the same time, thanks are due to the staff of the general office of our Department for their constant support and help.

Last but not the least, I am much grateful for my husband Fan Gao for his unconditional love and support, without whom this work would never be possible. In addition, the encouragement and support from my parents and parents-in-law have been of the utmost importance to me throughout the whole course of my pursuit of PhD study. I would like to thank my family from the deep bottom of my heart.

Contents

SUMMARY

With the fast advances in computing technologies, high-dimensional data have widely emerged in various areas, such as economics, bioscience, engineering, etc. In particular, when high-dimensional data sets are observed with the evolution of time, multivariate and high-dimensional time series modeling naturally attract massive research and empirical interest. However, high dimensionality poses numerous challenges and problems to modeling and implementations, due to the curse of dimensionality. With complex data features, we may encounter problems such as difficulties of identifying statistical models, infeasibility of numerical solutions and defective estimation results. Consequently, how to deal with these problems comes to be an essential step when we model and forecast high-dimensional data series.

Simultaneously, the existence of non-stationarity is also an inevitable issue to

handle in order to achieve desirable estimation and forecasting performance. Non-stationarity poses many challenges as well, not only for theoretical modeling but also for real time monitoring and forecasting. For instance, non-stationary modeling of financial returns has been discussed to be favourable in Mikosch and Stărică (1998) and Stărică and Granger (2005), among others. Under the explosion of a volatile market where stationary models are mis-specified, it is necessary to adopt non-stationary models to accurately capture the data dynamics. However, existing literature on the non-stationary issue for high-dimensional time series modeling is rather limited, compared to univariate cases. In this thesis, we are motivated to develop methods and models to analyze and forecast multivariate and high-dimensional time series under the existence of non-stationarity. The proposed models include factor model approach, adaptive multivariate approach and functional approach.

In the factor model approach, high dimensionality is reduced to a low-dimensional framework by applying functional principal component analysis (FPCA), with significant data information effectively preserved. A data-driven methodology is proposed to automatically select an optimal stationary time interval such that the accuracy of forecasting is improved, compared with a benchmark competitor. In the multivariate and functional approaches, the adaptive framework of a local univariate model is extended to both multivariate and functional domains respectively. In each of the two approaches, a simple underlying model structure is studied under the adaptive framework, which maintains the modeling parameter space at a reasonably low-dimensional level. Especially, in the functional approach, a consistent maximum likelihood (ML) estimator for functional autoregressive (FAR) model with nonzero mean function is derived. Theoretical properties of the proposed

adaptive estimate are also studied and proved in functional domain. Besides, with time-varying parameters, the proposed adaptive models can be safely applied to both stationary and non-stationary real world time series. Simulation study and real data applications are conducted for each of the proposed models. Reasonable and inspiring results are achieved in comparison with existing benchmark models.

List of Tables

List of Figures

CHAPTER 1

Introduction

Non-stationarity is an important and inevitable issue when real world time series data are considered. Its existence embeds in various areas and applications, such as economics, bioscience and engineering. Non-stationarity refers to the dynamics changes of data series as time evolves and it may result from the changes of statistical moments such as level and variance, as well as from the changes of the underlying modeling parameters used to describe the data series. The existence of non-stationarity poses many challenges, not only for theoretical modeling and statistical inferences but also for real time monitoring and forecasting, which

makes non-stationarity the very first problem to be solved in order to achieve desirable estimation and forecasting performance. Many works have been proposed to handle the non-stationary issue, see Tsay and Tiao (1984), Tsay (1984), Fan and Yao (2003) and references therein. Among them, most works are defined in a univariate framework, though some have been developed in multivariate or even high-dimensional scenarios. In this thesis, we are motivated to study adaptive modeling for multivariate and high-dimensional time series under the existence of non-stationarity. Three adaptive models are proposed, including factor model approach, adaptive multivariate approach and functional approach. Before we proceed to the proposed models, we will go through the existing literature briefly in the following sections.

1.1. Univariate non-stationary modeling

The existence of non-stationarity can usually be detected from the time evolution of raw data or its autocorrelation function (ACF) plot. As an illustration, on the left panel of Figure 1.1.1 we display the time series plot of monthly U.S. Treasury interest rates at 3-month maturity from January 1983 to December 2010. From the plot, an obvious downward trend is observed with the level and variation of interest rates changing over time. This example illustrates the existence

of non-stationarity in the real world data series, which is probably driven by the financial recessions starting in 1990, 2001 and 2007, see Chen and Niu (2014) for the data analysis of a similar period of U.S. Treasury data. The interest rate series, as shown in the figure, has a long memory with persistent sample autocorrelations. Such persistence feature is often observed in company with non-stationarity.

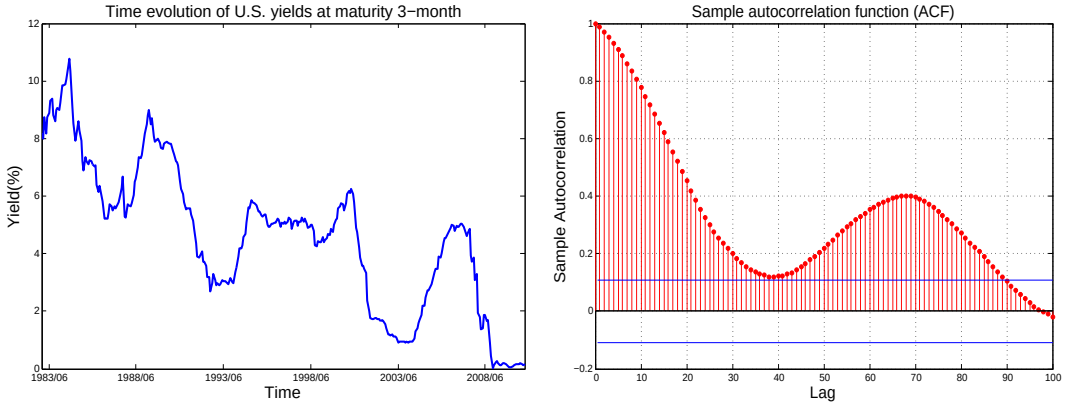

Figure 1.1.1 U.S. interest rates at maturity 3-month (left) and sample ACF plot (right) of the data. Data: monthly yield curves of U.S. Treasuries from January 1983 to December 2010.

The question on the true source of the persistence diagnosis, however, still remains to be answered. Diebold (1986) and Lamoureux and Lastrapes (1990) have noted that the presence of structural breaks may result in misleading inference on a long memory diagnosis. The theoretical results provided in Diebold and Inoue (2001) and Granger and Hyung (2004) further justify that this phenomenon can

also be spuriously generated by a short memory model with structural breaks or regime-shifts. More generally, Mikosch and Stărică (2004b) even argue independently that any particular model assumptions of non-stationarity, such as changes in the unconditional mean or variance, can lead to the diagnosis of long range dependencies. To address the persistence feature, modeling approaches can be broadly classified in two, the long memory approach and the short memory approach with structural changes. From the long memory view, the data generating processes are described by models with constant parameters and innovations with slowly or non-decaying effects, such as the fractionally integrated processes in Granger (1980), Granger and Joyeux (1980) and Hosking (1981). The short memory view considers persistence to be spuriously generated by changes in basic modeling parameters, such as heteroscedasticity, structural breaks or regime switching, see discussions in Diebold and Inoue (2001) and Granger and Hyung (2004). Technically, both the long memory view and the short memory view have merits in explaining persistence observed in the data. However, the short memory view often provides economic underpinnings to support various changes corresponding to policy shifts, regime transition and varying features of exogenous shocks, so on and so forth. In this thesis, we take the short memory view and will consider the persistence phenomenon as the consequence of non-stationary data structure.

The aforementioned academic findings have motivated the development of non-stationary short memory models, such as structural break detection methods (see e.g., Chen and Gupta, 1997; Mikosch and Stărică, 2004a; Liu and Maheu, 2008), time-varying coefficient models via Markov-Switching (see e.g., Hamilton and Susmel, 1994; So, Lam and Li, 1998) or via a smooth function of time or other transition variables (see e.g., Baillie and Morana, 2009; Scharth and Medeiros, 2009). Also, there is a large literature on GARCH models with explicit variation over time. They include the Spline-GARCH model of Engle and Rangel (2008), the GARCH-MIDAS model of Engle, Ghysels and Sohn (2008) and structural breaks models, see Andreou and Ghysels (2002) among others. While in the above mentioned works the model estimation is conducted using all the available information and under a parametric time-varying modeling structure, another class of local adaptive models has also been developed, see Belomestny and Spokoiny (2007), Čížek, Härdle and Spokoiny (2009) and Chen, Härdle and Pigorsch (2010). In these works, the time-dependent parameters are estimated under the assumption of local homogeneity. The local homogeneity assumes that there exists a local interval over which the data generating process can be well approximated by a stationary parametric model with constant parameters. In the modeling, the parameters of state variables are time-dependent without any explicit functional forms or any assumptions of change types, which makes the adaptive models flexible and universally suitable for both stationary and non-stationary time series.

Though the local adaptive models are desirable with flexibility and the ability of handling non-stationarity, they are developed in a univariate time series framework. The direct applications of the local models to multivariate time series involve difficulties, due to the curse of dimensionality. When the dimension of time series increases, challenging problems will be encountered, such as problems of model identification, estimator inefficiency and complexity in computations. These challenges would lead to low estimation accuracy or even misspecified modeling. In the following, we will proceed to the literature review on multivariate non-stationary time series analysis.

1.2. Multivariate non-stationary modeling

High-dimensional time series data have recently gained considerable popularity in areas of economics, biology, medical science and engineering. At the same time, high-dimensional models attract massive research and empirical interest whenever there arises the urgency of handling high-dimensional time series data. Similar to the univariate cases, non-stationarity never fails to make its appearance in modeling and forecasting high-dimensional time series. As an example, in Figure 1.2.1, we display the sample autocorrelations and cross-correlations of California hourly electricity log-prices for the whole sample from 5 July 1999 to 11 June 2000 in

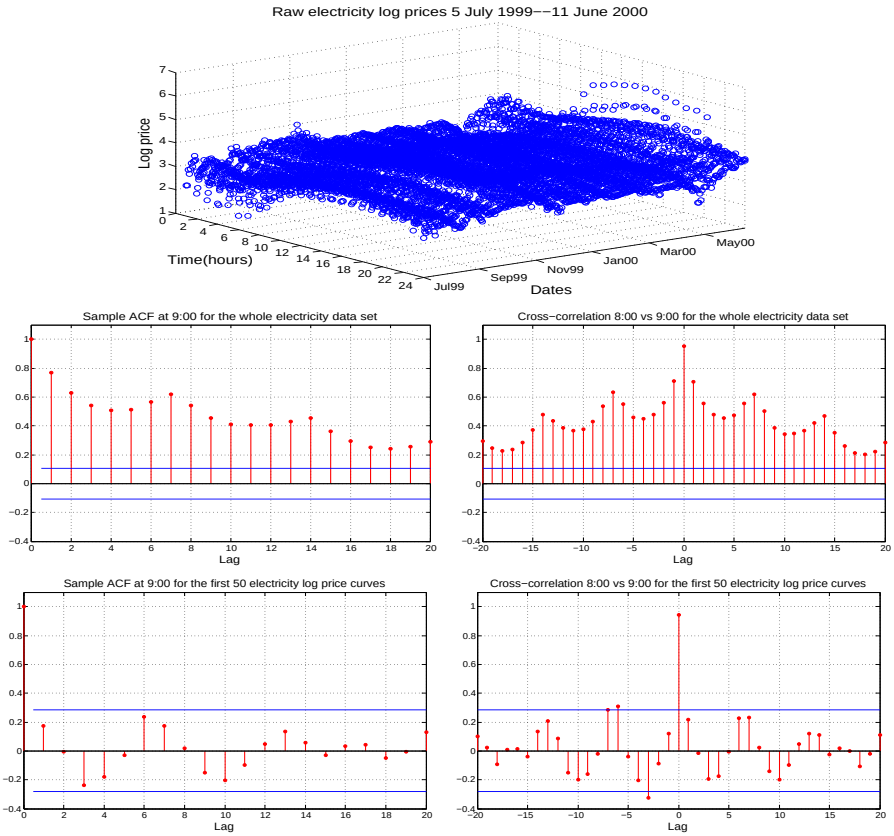

Figure 1.2.1 Sample autocorrelations of the log-prices at 9am and sample cross-correlations between 8am and 9am are displayed. Raw electricity log-prices are plotted at the top panel. The measures are computed using the whole sample from 5 July 1999 to 11 June 2000 in the middle panel. The bottom shows the respective sample autocorrelations and cross-correlations using a subsample from 5 July 1999 to 23 August 1999

the middle panel, and for a subsample from 5 July 1999 to 23 August 1999 at the bottom. The raw electricity log-prices data are plotted at the top of Figure

1.2.1. The serial dependence exits for both samples; however, the magnitude of dependence changes when the subsample is considered. This discrepancy regarding serial dependence indicates the possible existence of those not-that-sensational but eventually non-stationary events. There are a number of approaches to deal with non-stationarity in multivariate modeling; however in this thesis, we will focus our analysis and literature study on vector autoregressive (VAR) based models and factor models, motivated by the existence of serial dependence and their popularity in literature.

1.2.1 VAR based models

After Tong (1978) notes that the space in which the system is defined can be partitioned into two or more Euclidean spaces, with each separated Euclidean space representing a regime, there have been an increasing number of applications of threshold time series models. The threshold models capture the dynamic behaviour of data series such as periodic movements and regime changes, by switching to alternative AR or VAR models with different parameters among regimes. The switching mechanism is controlled by a threshold variable and associated threshold values, by which the conditions for different regimes are described. Although threshold models are assumed to be stationary, they can still be effectively applied to data series with regime switches or jump phenomena. For a thorough

discussion of threshold autoregressive models, we refer to Tong (1983) and the development and applications by Tong and Lim (1980), Tong (1987) and Tsay (1989), among others. It is worthwhile to note that Tsay (1998) proposes a threshold VAR (TVAR) model to generalize the model framework into multivariate settings. The TVAR type modeling has been widely applied to model business cycle effects, policy regime and transmission mechanism in developed economies, see Balke (2000), Atanasova (2003), Li and St-Amant (2010) and Afonso, Baxa and Slavik (2011). However, in threshold type models, the regimes are usually fixed which reduces the modeling flexibility. It is likely that failures of capturing and modeling unexpected jump phenomena may be encountered by using the pre-defined regimes. Another challenging issue for threshold modeling is the complexity of parameter estimation procedure. Parameters to be determined and estimated include the number of regimes, threshold variable, threshold values, and the modeling order and coefficients within each regime. In particular, as mentioned in Tsay (1998), there is often no best way to define the switching mechanism in real applications involving multiple time series. Therefore, a careful investigation is needed to determine an appropriate threshold variable or a switching mechanism for TVAR. In addition, the parameter estimation with more than two regimes has not been fully developed. Due to the additional computational complexity by extending to three or more regimes, most applications only focus on threshold modeling with two regimes, which makes the modeling set-up further restrictive.

Compared with TVAR models, time-varying VAR models are more flexible for studying the changing behaviour of economic systems. Since the late 1990s, time-varying parameters are considered in VAR modeling. Cogley and Sargent (2001) develop a VAR model with time-varying coefficients. They estimate a three-variable VAR model for post-war U.S. economic data with the variance term of structural shock restricted to be constant. To avoid modeling misspecification, Cogley and Sargent (2005) incorporate stochastic volatility into their time-varying VAR model, but leaving the simultaneous interactions among variables being time-invariant. Meanwhile, Primiceri (2005) proposes a VAR model with both time-varying coefficients and variance covariance matrix of innovations to study the changes in U.S. monetary policy over the post-war period. Unlike univariate time-varying coefficient models, most studies of time-varying VAR models assume random walk processes for the varying coefficients and stochastic volatility to avoid over-parameterization issue, due to the fact that a large number of parameters are introduced to the modeling by allowing time variation in parameters. Recently, there have been lots of applications and development of time-varying VAR models. We refer to Baumeister, Durinck and Peersman (2008) for the study on macroeconomics in European market, Galí and Gambetti (2009) for the study of sources of the Great Moderation and D'Agostino, Gambetti and Giannone (2013) for the investigation of forecasting performance of time-varying VAR models in comparison with other standard VAR approaches, among others.

It is noted that the incorporation of stochastic volatility brings an obstacle for parameter estimation mainly because the likelihood function now becomes intractable. To overcome this problem, a Bayesian approach using Markov Chain Monte Carlo (MCMC) methods is proposed and widely used for model estimation, see De Jong and Shephard (1995), Watanabe and Omori (2004) and Primiceri (2005). However, the estimation procedure is very tedious and complicated. Besides, when the time-varying VAR models are implemented in the Bayesian inference, the priors should be carefully chosen because there are many state variables and their processes are assumed to be non-stationary random walk processes, see Primiceri (2005) and Nakajima and Ginkō (2011) for the discussion and selection of priors to avoid undesired behaviour of time-varying parameters. Another challenging problem is the huge computational burden. In most existing time-varying VAR models, only a small number of lagged variables are considered; and it makes the computations highly demanding to incorporate more lagged or dependent variables when they are desired.

As mentioned, there are several drawbacks for both TVAR and time-varying VAR modeling. First of all, the flexibility of TVAR and time-varying VAR is achieved at the cost of substantially increasing the dimension of parameter space, which gives rise to the curse of dimensionality and further magnifies the computational burden. Furthermore, specific assumptions on the types and processes

of parameter changes are required. Also, the estimation is technically demanding and time consuming. To overcome these drawbacks, in Chapter 3, we extend a univariate adaptive process of local autoregressive model (LAR) to a local VAR (LVAR) framework. The generalized model can be applied to multiple time series for effective modeling and real-time applications in macroeconomics and finance. In contrast with TVAR and time-varying VAR modeling, LVAR is built on a simple underlying multivariate model, which is VAR of order 1. The adaptive procedure is designed for the selection of an optimal past interval for parameter estimation and forecasting, which saves the efforts of determining the lag order and thus maintains the dimension of parameter space at a lower level. As described in Chapter 3, the estimation procedure is relatively simple and the implementation of the methodology will not be hindered by any computational burden.

1.2.2 Factor models

When the number of observed time series increases, VAR based models cannot be practically applied because it is undesirable to include all the data series and expand the parameter space to a very high dimension. For such high-dimensional time series modeling, the challenges of high dimensionality in space and non-stationary dynamics in time are encountered at the same time. To mitigate the impact of high dimensionality, factor models are usually considered.

It is believed that only a small number of extracted factors are sufficient to explain and model the common behaviour of the original multiple time series, sometimes with an idiosyncratic component specific to each particular data series. Here we refer to Geweke (1977), Engle and Watson (1981), Peña and Box (1987), Forni, Hallin, Lippi and Reichlin (2000), Bai and Ng (2002), among others, for the existing literature. Based on the existing methods, non-stationary factor modeling is developed to deal with high dimensionality and non-stationarity simultaneously. Peña and Poncela (2004) analyze and study the structure and forecasting performance of a dynamic factor model, whose common factors are non-stationary. Del Negro and Otrok (2008) develop a dynamic factor model with time-varying factor loadings and stochastic volatility in both latent factors and idiosyncratic components. In addition, Pan and Yao (2008) propose a different method for estimating common factors, which are allowed to be non-stationary, by expanding the white noise space step by step. Eichler, Motta and Von Sachs (2011) propose a factor model with the factor loadings assumed to be functions of time and the idiosyncratic components allowed to be non-stationary with time-varying dynamics.

Alternatively, we can establish the modeling set-up in a functional framework. For instance in Chapter 2, the multiple time series of interest rates are treated and smoothed as a group of continuous functions of time to maturities, and statistical techniques to deal with functional data are applied for further analysis. Another

example can be found in Section 1.3 where the raw discrete electricity prices are considered as a collection of functions consisting of daily price curves. In this way, a parsimonious representation of data can be achieved. We refer to Section 1.3 for a systematic introduction of functional data and other advantages of modeling with them. In the proposed factor model, we adopt the functional view and propose a functional principal component analysis (FPCA) approach to identify the dominant factors, which are capable of representing the empirical data features. For a comprehensive review on theories and applications of FPCA, we refer to Ramsay and Silverman (2002), Ramsay and Silverman (2005) and Ferraty and Vieu (2006). The non-stationarity embedded in the original multiple data series can now be studied in a reduced low-dimensional space with only three extracted factors. From this functional point of view, we achieve a more parsimonious modeling framework and simpler estimation procedure which largely reduces the computational complexity, compared to the existing works. The proposed factor model is applied to U.S. and China Treasury interest rates; and the forecasting performance is compared with a benchmark competitor, the Dynamic Nelson-Siegel (DNS) model as studied in Diebold and Li (2006). Details of the modeling and empirical results can be referred to in Chapter 2.

1.3. Functional non-stationary modeling

The recent development of functional data analysis (FDA) sheds light on new approaches to model and analyze data sets of high dimensionality. FDA has become more and more popular in many scientific fields, where data are assumed to be generated by some underlying smooth functions. Apparently, for data of high dimensionality, it will be more appropriate to fit the data into functional framework instead of univariate or multivariate models. Even though such conventional statistical methods are available, FDA often provides more natural and parsimonious representation of data, and therefore leads to more accurate inference and prediction. Usually after smoothing the discrete observed data, we will obtain several curves over continuum such as time or ages. By interpreting data in this manner, the information between the observed data points generated by the same underlying smooth function can be readily recovered. Besides, assumptions such as stationarity for data points contained within one underlying curve, low dimensionality and equal spacing of sampling points are relaxed by considering each curve as a distinct observation or datum in FDA. Figure 1.3.1 displays an empirical example of the California electricity market. The discrete raw data points are the electricity log-prices observed 24 hours a day, 7 days a week from 5 July 1999 to 11 June 2000. This real data set is studied in Chapter 4.

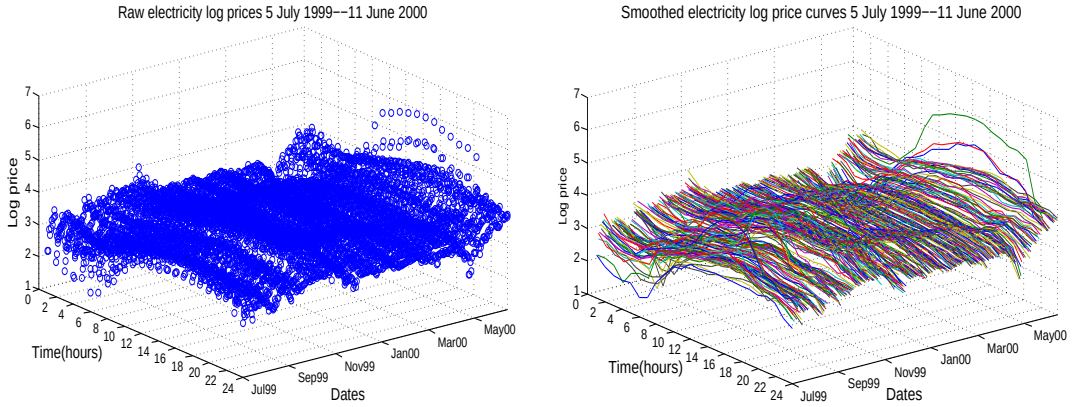

Figure 1.3.1 Left: Log-prices of the California electricity market for 24 hours a day, 7 days a week from 5 July 1999 to 11 June 2000. Right: Smoothed log-price curves of the California electricity market.

When each of the smooth curves is considered as a distinct datum for time points $t = 1, ..., n$, they form a time series with n functional objects. To study the autoregressive dependence structure of functional time series, one popular approach is functional autoregressive (FAR) modeling. Unlike univariate or multivariate time series models, we directly study the curves in the sense that the i-th curve is regressed on the previous p curves. The functional regressive model is defined to be an autoregressive Hilbertian process of order p (FAR(p)) when the generating process is assumed to be a sequence of random variables from Hilbert space. Extensive theoretical studies of FAR(p) can be found in Bosq (2000).

Before we proceed to propose the adaptive functional modeling, we will introduce the general FAR model and a brief literature review on estimating FAR operators. Let $\{X_1, ..., X_n\}$ denote a sample of random curves with sample size n and let $(\mathscr{H}, B_{\mathscr{H}})$ be a real separable Hilbert space endowed with its Borel σ-algebra $B_{\mathscr{H}}$ and with norm $\|\cdot\|$ induced from the inner product $\langle\cdot,\cdot\rangle$. The time series $\{X_t, t \in \mathbb{Z}\}$ is said to follow FAR(1) model if it satisfies $X_t - \mu = \rho(X_{t-1} - \mu) + \varepsilon_t$, where X_t take values in the Hilbert space $\mathscr{H}$ with mean function μ. The operator ρ is a bounded linear operator from $\mathscr{H}$ to $\mathscr{H}$ and the i.i.d. $\mathscr{H}$-valued sequence $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a strong $\mathscr{H}$-white noise with zero mean and $0 < E\|\varepsilon_t\|^2 < \infty$. In general, Bosq (2000) proves that, under some mild condition, a stationary solution exits for FAR(1) model.

The estimation of FAR operator ρ is firstly introduced in Bosq (1991) by approximating the solution of the functional Yule-Walker equation, namely $\Delta = \rho\Gamma$, where Γ is the covariance operator and Δ is the cross-covariance operator. After the empirical estimates of covariance operator and cross-covariance operator (i.e., $\hat{\Gamma}$ and $\hat{\Delta}$) are obtained, a naive estimator of ρ could be formulated as a composition of $\hat{\Delta}$ and $\hat{\Gamma}^{-1}$. However, $\hat{\Gamma}$ is not invertible in general. To solve this inverse problem, Bosq (1991) proposes to project data over a subspace $\mathscr{H}_{k_n}$ spanned by the first k_n eigen-functions of $\hat{\Gamma}$, from which an invertible operator in $\mathscr{H}_{k_n}$ can be obtained. For the consistency results and relevant proofs, see Bosq (2000). Pumo

(1998) further studies the estimation of ρ when the random functions take values in $C_{[0,1]}$. In addition, Guillas (2001) establishes the consistency results of a modified estimator of ρ, where a regulation parameter is incorporated into the estimate of Γ which provides a better control for the inverse of the covariance operator.

When $\mathscr{H}$ is the Sobolev space W^d, i.e., a function space in which any element function f defined on $[t_1, t_p]$ satisfies the condition that $f, f', \cdots, f^{(d-1)}$ are absolutely continuous and $f^{(d)} \in L^2([t_1, t_p])$, Besse and Cardot (1996) propose to project the data curves to a subspace of smooth functions and the proposed methodology accomplishes the dimension reduction and data smoothing simultaneously. The dimension of the subspace and the smoothing parameter are jointly determined by cross-validation. Further more, Besse, Cardot and Stephenson (2000) develop a kernel estimation method for covariance and cross-covariance operators to forecast climatic variations. Non-stationarity can be taken into consideration by defining a weighted estimator of the covariance operator, in the sense that the data curves closer to the last observation will be assigned a larger weight. It is worth to mention that Antoniadis and Sapatinas (2003) implement wavelet estimation methods to solve the inverse problem efficiently. Instead of projecting data to a subspace spanned by eigen-functions of the covariance operator, Kargin and Onatski (2008) develop a predictive factor decomposition to obtain an estimate of ρ by minimizing the expected squared norm of prediction error. The predictive factor approach

achieves better prediction performance when compared with the usual approach of projecting data to the subspace with PCA basis. Recently, Kokoszka and Zhang (2010) impose a positive threshold on the eigenvalues, which alleviates the inverse problem.

However, the aforementioned works assume stationarity in the functional data. Non-stationarity is likely to arise when we deal with functional time series, in the sense that the shape of the data curves or the structure of the FAR models may change with the evolution of time. Failure to take non-stationarity into consideration leads to erroneous modeling and inference. Theories and methodologies regarding non-stationarity in univariate cases have been well developed and some works have been successfully generalized to multivariate framework. However, little has been studied for non-stationary functional time series analysis, although we note that Besse et al. (2000) address the heteroscedasticity of covariance operator using a location-dependent weighting scheme and Horváth, Husková and Kokoszka (2010) propose a stability test of the FAR operator against a change point alternative. In this thesis, we propose a flexible modeling – Adaptive Functional Autoregressive (AFAR) model – for both stationary and non-stationary functional data, where the FAR operator, mean function and stochastic variations are allowed to be time-varying. In a stationary situation, where the parameters are constant, our developed consistent maximum likelihood (ML) estimators have closed forms.

In the non-stationary situation, we conduct estimation under local homogeneity which assumes the existence of a local interval over which the parameters are approximately constant. In other words, the underlying data generating process is approximated by a stationary FAR model with constant parameters. The local interval is identified in a sequential testing procedure, and a set of data-driven critical values is calibrated and used to measure the significance of the divergence between the time-varying AFAR model and the constant FAR model.

1.4. Proposed methods and contributions

We propose and study three models to deal with the non-stationary issue for high-dimensional time series with applications in economics and financial market. We extend the univariate adaptive modeling to both multivariate and functional domains with some theoretical results. In addition, we develop consistent ML estimators with closed-forms in FAR modeling. The consistency results and the theoretical properties of adaptive estimates are studied and proved in the functional approach. The proposed adaptive models are applied to the multivariate and high-dimensional time series data of U.S. and China Treasuries and California electricity prices. The modeling and forecasting performance are generally improved and more accurate prediction under the existence of non-stationarity is achieved.

Firstly, we propose a factor model to adaptively forecast yield curves. The FPCA method is used to extract dominant factors to represent the empirical features of yield curves. After the temporal dynamics are pushed into the factor loadings, the univariate adaptive autoregressive approach LAR is introduced to handle the non-stationarity inherent in the data dynamics. It is necessary to emphasize that the main contribution of the factor approach is not developing a new estimation method or related theoretical results. Instead, a data-driven technology is proposed to automatically select a reliable stationary time interval such that the accuracy of yield curve forecasting is improved. Though no further theory is developed, we study the applications of this local modeling in the prediction for U.S. and China yield curves. Compared to its natural competitor, the DNS model, our proposed model provides reasonable performance in both simulation study and empirical analysis.

Secondly, we develop a multivariate non-stationary model, referred to as LVAR in this thesis. In the multivariate model, the adaptive procedure is generalized from univariate underlying data generating processes to multivariate cases. An interval of local homogeneity is carefully selected at any particular time point. With this flexibility on interval selection, the LVAR model provides stable performance both in a simulated homogeneous situation and under regime shift scenarios. Compared

with the existing literature, LVAR modeling has a simple underlying data generating process, which maintains the dimension of parameter space at a lower level. Moreover, the computational complexity is reduced due to its relatively simpler estimation procedure.

In addition, we work on an AFAR model to handle non-stationarity for functional time series. We derive new consistent ML estimators for the general FAR model with a nonzero mean function. This is particularly important for non-stationary functional data, where we are unable to detect structural changes before estimating the model and thus we could not simply demean or adopt a zero mean function assumption as universally assumed in stationary scenarios. In our study, the likelihood function is re-defined, which is different from the existing work in Mourid and Bensmain (2006). Our definition allows to derive the ML estimators not only for the FAR operator but also for the stochastic variations. With this newly defined density function and likelihood function, the consistency results of the ML estimators are proved. Details can be found in Chapter 4. Moreover, with time-varying parameters, the AFAR modeling is flexible and can be safely applied to both stationary and non-stationary functional data. Our estimation is conducted on local intervals with time-varying lengths for each particular time point, instead of a fixed bandwidth as in the rolling window technique and the similar one in Besse et al. (2000). The theoretical results for the final adaptive estimates

are also proved and discussed in Chapter 4.

All the three works are proposed to model and predict high-dimensional time series under the existence of non-stationarity. Inspiring results are obtained, compared with the existing methods. The LVAR modeling can be directly used to work on multivariate time series. However, the forecast accuracy might be low when LVAR is applied to high-dimensional data series, due to the joint impact of curse of dimensionality and non-stationarity. For both factor model and AFAR model, the data are considered as functional data and smoothed data curves are studied for further modeling and analysis. Besides, dimension reduction techniques are utilized in both models but with different focuses. In the factor model, FPCA is used to reduce the dimensionality of data space to the number of dominant factors. In this way, the infinite-dimensional functional data can be represented by only three extracted factors or basis functions. The modeling estimation is based on the three extracted factors, which are studied and predicted for the forecast of the functional data. While in AFAR model, we reduce the dimensionality of parameter space by using the method of sieves. Instead of handling an infinite parameter space, the estimation procedure is performed within a finite subset of the parameter space. The subset is called sieve and its dimension is allowed to increase with the sample size. As we have proved, the sieve estimator is a consistent estimator, and it is used for future prediction.

The remainder of the thesis is structured as follows. In Chapter 2, we introduce the proposed factor model by using FPCA and the adaptive modeling set-up is based on the univariate time series for each of the extracted factors. The univariate adaptive framework is generalized to multivariate data in Chapter 3 and functional data which is infinite-dimensional in Chapter 4. Details of the three models and all the corresponding numerical studies will be discussed in the following chapters.

CHAPTER 2

Factor model with FPCA

In Chapter 2 and 3, we will study and investigate the proposed models with applications of modeling and forecasting yield curves. In this chapter, before we introduce the proposed factor model, a brief literature review of modeling yield curves by existing factor models is studied. As is known, yield curve plays an important role in economies, which depicts interest rates against maturities. However, as the number of interest rate maturities increases, it poses difficulties in accurately estimating the parameters when some parametric models are applied, above all due to the curse of dimensionality, see Härdle, Müller, Sperlich and Werwatz (2004). This inaccuracy would create non-ignorable risk in applications. To

deal with such a feature, factor models have been developed to analyze interest rates or term structure, see e.g., Chen (1996), Schaefer and Schwartz (1984) and Hull and White (1994). By extracting a small number of dominant factors, the problem is converted to a low-dimensional one, see Vetzal (1994). Among others, Nelson-Siegel (NS) model (e.g., Nelson and Siegel, 1987; Svensson, 1995) and Dynamic Nelson-Siegel (DNS) model (Diebold and Li, 2006) are by far the most popular models, which obtains 3 factors based on exponential factor loadings. In Figure 2.0.1 we depict the NS factor loadings which correspond to the factors representing the level, slope and curvature of China yield curves.

In spite of the popularity of NS model, it is natural to ask whether the NS exponential factor loadings are universally appropriate for any kind of yield curves. In Figure 2.0.1, we also display 3 empirical factor loadings for the monthly yield curves of China Treasuries. These curves are obtained by using FPCA, which is a data-driven method capable of extracting factors consistent to the empirical features of the data. More specifically, the respective FPCA factors account for above 99% variations of the raw data and can well represent the level, slope and curvature of the yield curves. The details of the FPCA method can be found in Section 2.2.1. It is interesting to note that the shape of the empirical curvature factor loadings deviates much from the conventional NS curve, with an obvious double-humped shape peaking not only around the medium maturity but also

around the long maturity. This possibly refers to the unique sovereign credit risks or central bank regulations.

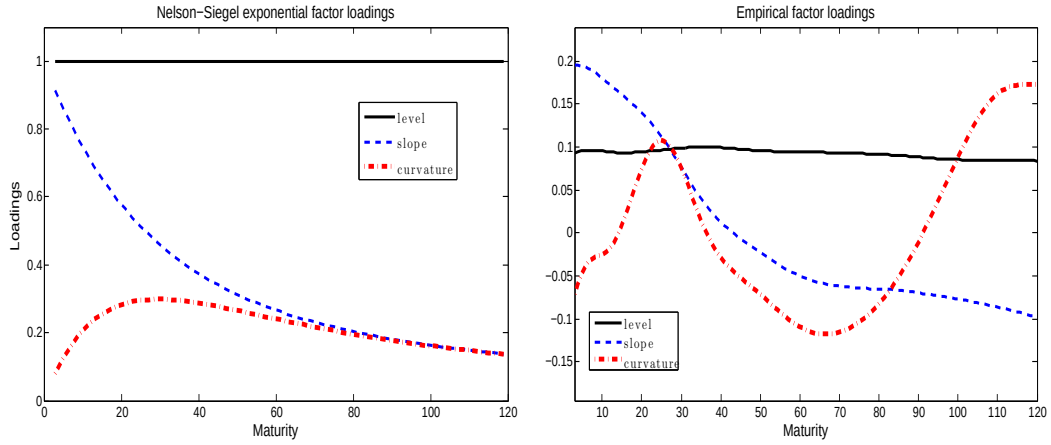

Figure 2.0.1 The empirical factor loadings of the China yield curves (right) and the NS exponential loadings (left). In the NS framework: the level loading is 1; the slope loading is $(1-e^{-\lambda_t\tau})/\lambda_t\tau$ and the curvature loading is $(1-e^{-\lambda_t\tau})/\lambda_t\tau - e^{-\lambda_t\tau}$, with $\lambda_t = 0.0609$ and τ denoting the time to maturities. Data: monthly yield curves of China Treasuries from March 2003 to October 2011, Datastream.

The recent development of FDA sheds light on new approaches to obtain factors of yield curves. In particular, FPCA provides a neat and efficient methodology to extract dominant factors. By considering yield curves as functional data, with each yield curve naturally representing a function of maturities, FPCA method extracts factors that explain the maximal variation of the curves via orthogonal decomposition. Compared to NS models, FPCA method is appealing as it takes into account

the natural functional feature of yield curves and seeks factors according to data's empirical dependence structure. This method also helps to identify unique features and can be safely used for any kinds of yield curves. It is worth mentioning that FPCA is sufficiently different from the multivariate PCA method. The former handles data as curves that correspond to infinitely many maturities, while the latter considers data as discrete points with multivariate but finite maturities, see Müller (2005).

Moreover, the existing forecast models are largely based on the stationarity assumption of the factors and the model set-up is rather restrictive. In the widely-used DNS model, AR(1) specification for each of the three factors is revealed to be superior to many competitors, including the random walk model, slope regression, Fama-Bliss forward rate regression (Fama and Bliss, 1987), affine model (see e.g., Duffie and Kan, 1996; Egorov, Li and Ng, 2011), VAR models and error correction models (Engle and Granger, 1987). The DNS model however ignores the structure changes and regime shifts that exist especially at the time when markets are volatile. Figure 2.0.2 for example displays the time evolution of the resulting NS level factor which is extracted based on the monthly yield curves of U.S. Treasuries, as in Diebold and Li (2006). The sample autocorrelations are persistent and this type of persistence has been addressed in the literature of non-stationarity of interest rates. Hall, Anderson and Granger (1992) show that interest rates at different

maturities are co-integrated and driven by unit-root non-stationary factors. Bansal and Zhou (2002) develop a model where the short rate and the market price of risks are subject to regime shifts. Guidolin and Timmermann (2009) propose a regime switching VAR model for an aggregated forecast of the U.S. short-term interest rates, in which the aggregation weights shift between regimes.

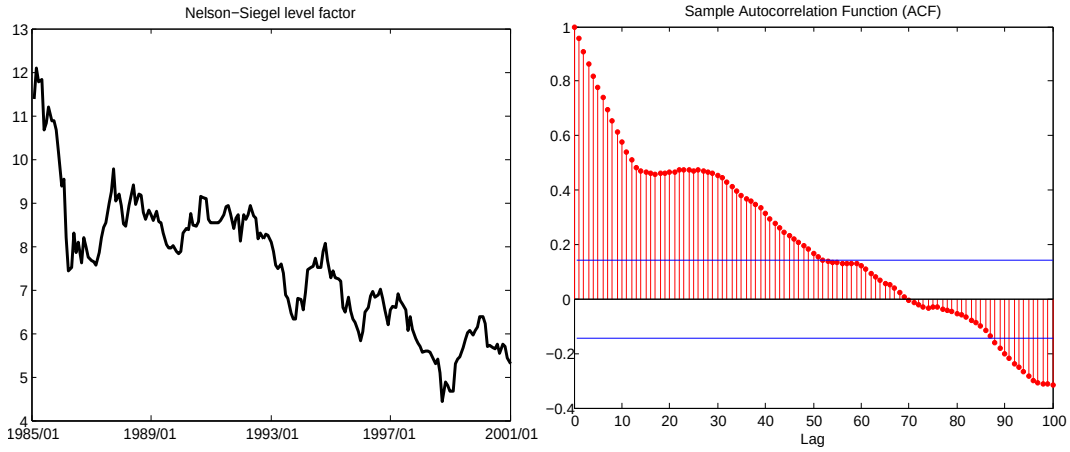

Figure 2.0.2 The level factor based on the Nelson-Siegel exponential basis (left) and its sample ACF plot (right). Data: monthly yield curves of U.S. Treasuries from January 1985 to December 2000, see also Diebold and Li (2006).

In this chapter, a factor model with the combination of FPCA and univariate adaptive modeling LAR is proposed and applied to forecast yield curves. Now we will introduce further about the proposed model and the empirical study. The content and materials presented in this chapter have been published in Chen and Li (2011).

2.1. Smoothing of the data

Two data sets are considered, the U.S. Treasuries and the China Treasuries. The U.S. data consist of the end-of-month price quotes (bid-ask average) for U.S. Treasuries, spanning from January 1985 to December 2000. There are 192 monthly interest rates at 17 maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. The data has been studied in Diebold and Li (2006). The second data set contains the end-of-month price quotes for China Treasuries from March 2003 to October 2011. There are 104 monthly interest rates at 11 maturities of 3, 6, 12, 24, 36, 48, 60, 84, 96, 108 and 120 months.

The underlying yield curve, as a function of maturities, is not directly observable. Parametric and nonparametric methods have been developed to estimate yield curve based on the available discrete interest rates, among which nonparametric methods often provide a better fit. Polynomial splines are the most popular nonparametric techniques used for estimating yield curve, see McCulloch (1975), McCulloch (1971), Schaefer (1973), Vasicek and Fong (1982) and Shea (1985). The methods are sensitive to the selection of smoothing parameters such as knots in splines. For the selection, we refer to Jarrow, Ruppert and Yu (2004) and Fernández-Rodríguez (2006) among many others.

In our study, we use B-splines smoothing technique to obtain the functional data. Let $\mathbf{X}_t(\tau_1, \cdots, \tau_q)$ be the discrete interest rates at time point $t = 1, \cdots, T$ that contain q maturities and $X_t(\tau)$ denote the yield curve, a function of maturity $\tau \in \mathbb{R}$. The yield curves are estimated as:

$$X_t(\tau) = \sum_{k=1}^{K} c_{tk}\phi_k(\tau) = c_{t1}\phi_1(\tau) + \cdots + c_{tK}\phi_K(\tau),$$

where $\phi_1(\tau), \cdots, \phi_K(\tau)$ are K basis functions. We refer to Ramsay and Silverman (2005) for a justification of this selection. The coefficients $c_{t1}, \cdots, c_{tK}$ are estimated by minimizing the penalized sum of squared errors:

$$\text{PENSSE}_\lambda = \|\mathbf{X}_t(\tau_1, \cdots, \tau_q) - X_t(\Lambda)\|^2 + \lambda \int [D^2 X_t(\tau)]^2 d\tau,$$

where $X_t(\Lambda)$ contains the function values of the smoothed curve at the discrete maturities $\Lambda = (\tau_1, \cdots, \tau_q)$, $D^2 X_t(\tau)$ is the second derivative function of $X_t(\tau)$ and the parameter λ is a smoothing parameter that controls the smoothness of the estimated curve. $\|\cdot\|$ is used to denote the L^2 norm. In Figure 2.1.1, the smoothed yield curves are displayed in 3D view. In the plot, the movement of yield curves as time goes by can be clearly captured. The percentage errors across all maturities are calculated. By accessing the fitted percentage errors, we will easily see that the smoothed curves via B-splines serve as a reasonable representation of the underlying yield curves because the average percentage errors among all curves are 1.5309×10^{-5} and 6.5702×10^{-6} for U.S. and China data respectively. The smooth curves are the functional data considered in the following sections.

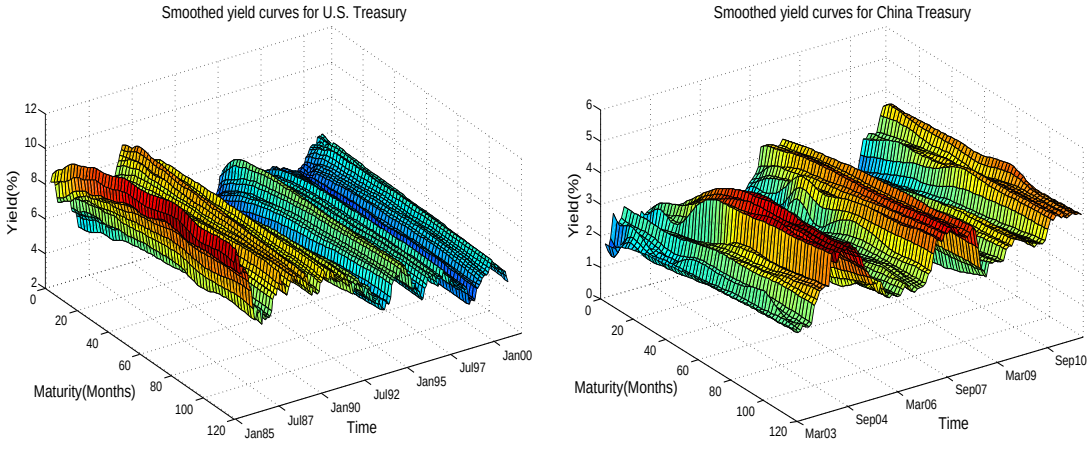

Figure 2.1.1 The estimated yield curves for U.S. Treasuries (left) and China Treasuries (right) via B-splines.

2.2. Method

The FPCA method projects yield curves into the directions which explain the largest variations of data. As an illustration, the empirical loadings in Figure 2.0.1 are obtained by FPCA. It is worth noting that, via a linear transformation, any form of non-stationarity in yield curves would be attributed to the processes of the resulting factors. A time-varying autoregressive (AR) model is used to model and forecast each of the factors, where the parameters are estimated under the assumption of local homogeneity. That is, for any particular time point, there exists a past time interval over which the data can be well represented by an AR process with constant parameters. It is analogous to the rolling window technique

(with fixed window size), though in the local model, the window size changes over time t. The time intervals are identified in a data-driven way.

2.2.1 Extracting factors via FPCA

The yield curve is denoted as $X_t(\tau)$ at time point $t \in [1, T]$, which is a function of time to maturities $\tau \in \mathbb{R}$. Without loss of generality, the yield curves are assumed to be demeaned, i.e., $E[X_t(\tau)] = 0$. FPCA is used to extract factors:

$$f_t = \int \xi(\tau) X_t(\tau) \mathrm{d}\tau, \tag{2.1}$$

where f_t denotes the factor and ξ the corresponding factor loadings. The solution is obtained by maximizing the variation of the respective factors f_t:

$$\max \frac{1}{T} \sum_t f_t^2 = \frac{1}{T} \sum_t \left\{ \int \xi(\tau) X_t(\tau) \mathrm{d}\tau \right\}^2, \quad \text{subject to } \int \xi^2(\tau) \mathrm{d}\tau = 1.$$

Here the factor loadings follow the unit-norm condition to guarantee a unique solution. To obtain the other factors, the factor loadings are assumed to be orthogonal with $\int \xi_k(\tau) \xi_m(\tau) \mathrm{d}\tau = 0$, for all $k \neq m$ and $k, m \leq p$, where p is the number of selected factors.

The factor loadings can be estimated by solving eigen-decomposition. Define the covariance of yield curves as

$$v(\tau, s) = \frac{1}{T} \sum_t X_t(\tau) X_t(s),$$

where $\tau \in \mathbb{R}$ and $s \in \mathbb{R}$. Figure 2.2.1 displays the sample covariance surfaces of the U.S. and China yield curves respectively. Clearly, covariance becomes larger when maturities are closer, and decays when maturities are farther apart. The factor

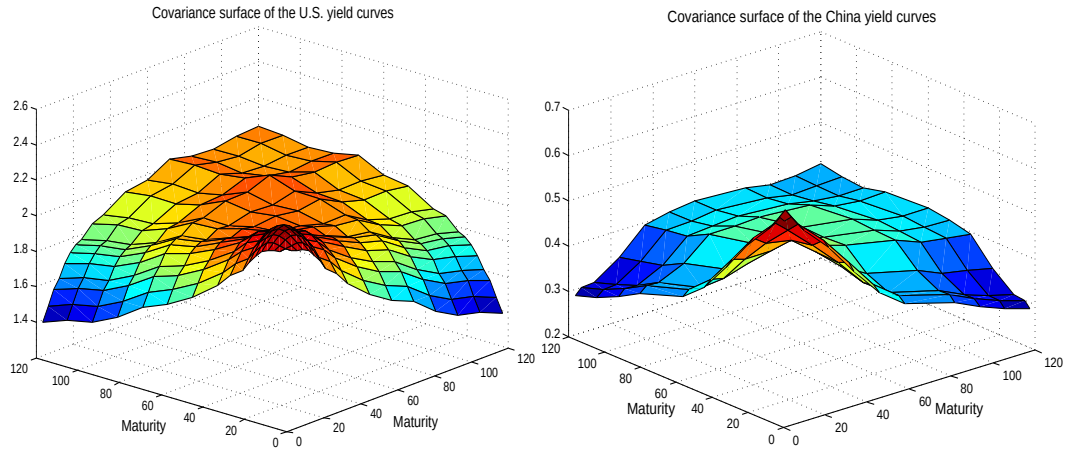

Figure 2.2.1 The sample covariance surfaces of the yield curves of U.S. Treasuries from January 1985 to December 2000 (left) and of China Treasuries from Mar 2003 to Oct 2011 (right).

loadings $\xi(\tau)$ is actually an eigenfunction of the covariance and

$$\int v(\tau, s)\xi(\tau)\mathrm{d}\tau = \alpha\xi(s), \tag{2.2}$$

where α denotes the eigenvalue.

Ramsay and Silverman (2005) propose a solution of Equation (2.2). By expressing the functional data $\mathbf{X}(\tau)$ with K basis functions $\Phi(\tau) = \left[\phi_1(\tau), \cdots, \phi_K(\tau)\right]^\top$,

$$\mathbf{X}(\tau) = \begin{bmatrix} X_1(\tau) \\ X_2(\tau) \\ \vdots \\ X_T(\tau) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{K} c_{1k}\phi_k(\tau) \\ \sum_{k=1}^{K} c_{2k}\phi_k(\tau) \\ \vdots \\ \sum_{k=1}^{K} c_{Tk}\phi_k(\tau) \end{bmatrix} = \mathbf{C}\Phi(\tau), \tag{2.3}$$

where $\mathbf{C}$ is a $T \times K$ matrix of coefficients, we can formulate the weight function $\xi(\tau)$ in a basis expansion with the same basis functions, but different coefficients $\mathbf{b} = (b_1, \cdots, b_K)^\top$:

$$\xi(\tau) = \Phi^\top(\tau)\mathbf{b}. \tag{2.4}$$

Substituting the expansions (2.3) and (2.4) into (2.2), we have:

$$\frac{1}{T}\Phi^\top(s)\mathbf{C}^\top\mathbf{C}\int \Phi(\tau)\Phi^\top(\tau)\mathrm{d}\tau\mathbf{b} = \alpha\Phi^\top(s)\mathbf{b}.$$

Note that this equation applies for all values of s and hence we can drop $\Phi^\top(s)$. By defining a $K \times K$ matrix $\mathbf{W} = \int \Phi(\tau)\Phi^\top(\tau)\mathrm{d}\tau$ and a vector $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{b}$, we have:

$$\frac{1}{T}\mathbf{W}^{1/2}\mathbf{C}^\top\mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \alpha\mathbf{u},$$

and $\mathbf{b_m^\top W b_m} = 1$, $\mathbf{b_k^\top W b_m} = 0$ for $k \neq m$ and $k, m \leq p$, where $\mathbf{b_m}$ corresponds to the coefficient vector of the m-th eigenfunction $\xi_m(\tau)$. The eigenvalue α and eigenvector $\mathbf{u}$ are solvable for the $K \times K$ matrix $\frac{1}{T}\mathbf{W}^{1/2}\mathbf{C}^\top\mathbf{C}\mathbf{W}^{1/2}$. Consequently, we have:

$$\xi(\tau) = \Phi^\top(\tau)\mathbf{b}, \quad \mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}.$$

An interesting question is how many factors are needed in order to capture the dynamics of yield curves. In other words, how to select the number of factors p. Litterman and Scheinkman (1991) show that three factors are necessary for U.S. yield curves, with the factors representing the level, slope and curvature; while Cochrane and Piazzesi (2005) and Dai, Singleton and Yang (2004) show that up to five factors should be considered for U.S. government bonds. Egorov et al. (2011) find four factors (two common and two local) for the joint yield curves of U.S. and euro interest rates. In addition to these selections, FPCA provides a natural quantitative selection criterion. According to (2.2), the i-th largest eigenvalue α_i associates to the variation direction of yield curves defined by $\xi_i(\tau)$, $i = 1, \cdots, K$. Therefore, the cumulative proportion of variation explained by all the p factors is:

$$\frac{\alpha_1 + \cdots + \alpha_p}{\alpha_1 + \cdots + \alpha_K}.$$

In our study, we select in principal $p = 3$ factors and this selection is motivated by the economic meanings of the factors. It is however suggestive to select more factors if the cumulated sum of variation explained is below 99%.

2.2.2 Fitting a LAR model to the factors

In the following, we briefly introduce the local model for estimating and forecasting the FPCA factors. Since the model structure and estimation procedure are

the same for all the p selected factors, we drop the subscript j of the factors f_{jt} with $j = 1, \cdots, p$ for notational simplification. Therefore, without loss of generality, $\{f_t\}$ now stands for the univariate time series of any of the p factors. For each $\{f_t\}$ taking values in $\mathbb{R}$, the LAR model of order 1 or LAR(1) model, is defined through a time-varying parameter set $\theta_t = (\theta_{0t}, \theta_{1t}, \sigma_t)^\top$ such that

$$f_t = \theta_{0t} + \theta_{1t} f_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_t^2).$$

The estimation of the parameters is conducted under the assumption of local homogeneity at each time point t. Under local homogeneity, the parameter θ_t is assumed not to deviate much from a constant in a local interval $I_t = [t - m_t, t)$, and hence the data are (approximately) stationary. The local maximum likelihood estimator $\hat{\theta}_t$ is defined over the local interval in the following way:

$$\begin{aligned} \hat{\theta}_t &= \text{argmax}_{\theta \in \Theta} L(I_t, \theta) \\ &= \text{argmax}\Big\{ - m_t \log\sigma_t - \frac{1}{2\sigma_t^2} \sum_{s=t-m_t}^{t-1} (f_s - \theta_0 - \theta_1 f_{s-1})^2 \Big\}, \end{aligned}$$

where Θ denotes the parameter space and $L(I_t, \theta)$ is the local log-likelihood function. It is worth noting that the local estimation method is different from the rolling window technique with a globally constant window size. Here the local window size m_t is time dependent and in practice unknown. The question is of course how to select the local interval or the value of m_t. Generally speaking, the optimal selection would be the longest interval where the local homogeneity assumption holds, that is, the time series of the factor can be well described by a

model with constant parameters. It is noted that long intervals provide accurate estimates with low variation; however the local homogeneity assumption is likely to be violated, as the interval length further increases. This violation will possibly lead to a large modeling bias. Therefore, the optimal selection is designed to balance the modeling bias and variance tradeoff.

We employ an automatic procedure to select the intervals. At every time point, it starts with a small sample size m_0 that defines a local interval $I_t^{(0)} = [t - m_0, t)$. The value of m_0 is small enough to ensure the homogeneity, where a conventional AR(1) model has a reasonable fit. Iteratively, we increase the sample size to m_k, with $k > 0$ and $m_k > m_{k-1}$, which defines a longer interval $I_t^{(k)} = [t - m_k, t)$. At this moment, if $I_t^{(k-1)} = [t - m_{k-1}, t)$ has been accepted as an interval of homogeneity, we can then check the deviation of estimation on $I_t^{(k)}$ from homogeneity by a likelihood ratio:

$$\begin{aligned} T_t^{(k)} &= L(I_t^{(k)}, \hat{\theta}_t^{(k-1)}) - L(I_t^{(k)}, \hat{\theta}_t^{(k)}), \\ \text{with } \hat{\theta}_t^{(k-1)} &= \text{argmax}_{\theta \in \Theta} L(I_t^{(k-1)}, \theta), \\ \hat{\theta}_t^{(k)} &= \text{argmax}_{\theta \in \Theta} L(I_t^{(k)}, \theta), \end{aligned}$$

where $\hat{\theta}_t^{(k-1)}$ is an accepted estimate under local homogeneity over the interval $I_t^{(k-1)}$. If the difference is small, which indicates the larger data sample displays similar patterns as the smaller one, we would accept the longer interval for an improved accuracy of estimation. On the contrary, if the difference is large, it

implies that the modeling changes have occurred. In this case, we terminate the procedure to avoid substantial modeling bias. The lastly accepted interval would be the optimal choice. One continues this way until a change is suspected or the possibly longest interval is screened. Chen et al. (2010) have developed a test to measure the significance level of the difference. We refer to the work for more details.

2.3. Simulation

The proposed method includes two parts, extracting factors via FPCA and forecasting in the LAR framework. In this section we investigate the performance of the proposed method in a practical simulation study. More specifically, we study the describability of FPCA under two scenarios: 1) U.S. scenario, where the yield curves are driven by the NS exponential factor loadings and 2) China scenario, where the true underlying process is generated by the empirical loadings obtained via FPCA. Both the FPCA and DNS methods are applied. The root mean square error (RMSE) is used to assess the estimation accuracy between the actual (generated) and estimated yield curves.

In the U.S. scenario, yield curves are generated by following the DNS modeling

framework:

$$X_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) + \varepsilon_t(\tau), \tag{2.5}$$

where $\lambda_t = 0.0609$ which maximizes the curvature loading at a medium maturity of 30 months, see Diebold and Li (2006). The factors β_{jt} with $j = 1, 2, 3$ are assumed to be known and they are obtained as the ordinary least squares (OLS) estimates by fitting DNS model to the monthly U.S. Treasuries from January 1985 to December 2000. In the simulation, we generate 192 yield curves, each containing 17 yields at maturities from 3 months to 10 years for $t = 1, \cdots, 192$. The stochastic innovations $\varepsilon_t(\tau)$ are i.i.d normal random variables, with mean and standard deviations calculated from the fitted errors of the DNS modeling of U.S. Treasuries data. We repeat the generation for 500 times.

Analogously, the China scenario is designed based on the China Treasuries from March 2003 to October 2011. The data consist of 104 monthly yield curves corresponding to 11 maturities from 3 months to 10 years. We use FPCA to estimate the values of the FPCA factors and the empirical factor loadings by the following:

$$X_t(\tau) = f_{1t}\xi_1(\tau) + f_{2t}\xi_2(\tau) + f_{3t}\xi_3(\tau) + \varepsilon_t(\tau),$$

where f_{jt} refers to the resulting factor and $\xi_j(\tau)$ the empirical factor loadings, with $j = 1, 2, 3$. Again, the error term $\varepsilon_t(\tau)$ is assumed to be normally distributed with mean and standard deviation calculated from the fitted errors of the FPCA

modeling with China Treasuries data. The resulting factor loadings have been depicted in Figure 2.0.1. In the simulation study, the estimated FPCA factors and the associated factor loadings are used to generate 104 curves, each including 11 interest rates. The generation is also repeated for 500 times.

Both the FPCA and DNS methods are used to extract factors for the generated data in the two scenarios. In the DNS modeling, three factors are extracted. In FPCA, we also select three dominant factors, as they explain more than 99% of the total variation for each case. The average values of the accumulated proportion of the explained variance for the three factors are $94.2\%, 99.7\%, 99.9\%$ in the U.S. scenario and $89.3\%, 98.8\%, 99.3\%$ in the China scenario. The NS exponential factor loadings are fixed, which however fail to represent the underlying data structure in the China scenario while the FPCA loadings differ among data. This data-driven method adapts according to the actual dependence structure of data. In the U.S. scenario, the FPCA factor loadings well replicate the exponential curves, though with different magnitudes. The scaling deviation is due to the demeaning process and it has no impact on yield curve forecast as the mean process is added back. In the China scenario, the empirical factor loadings are good proxies of the underlying curves. As an illustration, We depict the FPCA factor loadings of one randomly selected data set for each of the two scenarios, see Figure 2.3.1. The results for the other generated data sets behave in a similar way, which are omitted here.

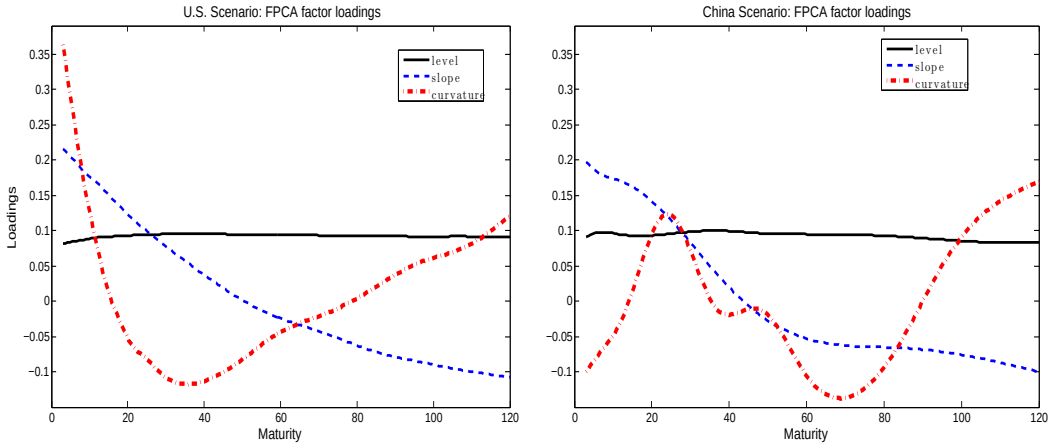

Figure 2.3.1 One realization of the FPCA factor loadings for both simulation scenarios. In the DNS or U.S. scenario, the resulting factor loadings well represent the underlying NS exponential curves (left). In the FPCA or China scenario, the resulting factor loadings are good proxies of the underlying curves, too (right).

The RMSE values between the fitted yield curves and the generated yield curves are computed. The smaller are the values, the higher the accuracy. Table 2.1 reports the average values of the RMSE at various maturities. It reveals that the FPCA method is indeed superior to the DNS model for most maturities. In the China scenario, FPCA works well for maturities from 1-year to 10-year. Even in the U.S. scenario, FPCA performs better for 15 maturities, except the 3-month and 10-year maturities. The main reason of the relatively worse performance at the shortest horizon 3-month and sometimes the longest horizon 120-month, is possibly due to the boundary problems when B-splines smoothing technique is applied. In the relative term, the FPCA method improves the estimation accuracy up to 16%

in the U.S. scenario and 29% in the China scenario. The results show that FPCA performs better in both scenarios. It not only improves estimation accuracy but also effectively captures the underlying pattern of the data, whereas the alternative model encounters mis-specification in the China scenario.

2.4. Real Data Analysis

In this section, we implement the FPCA-LAR method to forecast the yield curves of U.S. Treasuries and China Treasuries. We investigate the performance of the adaptive method compared to its natural competitor, the DNS model with AR(1) specification (Diebold and Li, 2006). Does the proposed forecast method benefit from using the data-driven factor extraction and the adaptive modeling? To address this question, we assess and compare the out-of-sample forecast accuracy of the methods. Let us now describe how we forecast yield curves using the proposed method and its competitor.

(1) FPCA-LAR: at each time point t, we apply FPCA to extract factors and fit an LAR(1) model to each factor over a selected optimal local interval. The $h-$step ahead forecasts of the factors as well as the yield curves are:

$$\begin{aligned} \hat{X}_{t+h}(\tau) &= \sum_{j=1}^{p} \hat{f}_{j,t+h}\xi_j(\tau), \\ \text{with } \hat{f}_{j,t+h} &= \hat{\theta}_{0jt} + \hat{\theta}_{1jt} f_{j,t}, \end{aligned}$$

	U.S. scenario		China scenario	
maturity	DNS	FPCA	DNS	FPCA
3	0.05	0.08	0.08	0.12
6	0.05	**0.04**	0.09	0.10
9	0.06	**0.05**	–	–
12	0.07	**0.07**	0.09	**0.06**
15	0.08	**0.05**	–	–
18	0.06	**0.04**	–	–
21	0.04	**0.04**	–	–
24	0.05	**0.04**	0.11	**0.06**
30	0.04	**0.04**	–	–
36	0.06	**0.04**	0.08	**0.08**
48	0.06	**0.05**	0.09	**0.08**
60	0.08	**0.05**	0.10	**0.03**
72	0.07	**0.07**	–	–
84	0.06	**0.05**	0.07	**0.04**
96	0.05	**0.04**	0.06	**0.06**
108	0.06	**0.04**	0.06	**0.03**
120	0.06	0.07	0.08	**0.06**

Table 2.1 Simulation results. The average values of RMSE between the fitted and actual (generated) interest rates are reported for the two scenarios. Both the DNS and FPCA methods are used in each scenario. The results with smaller errors are marked in bold to highlight better accuracy.

where $\hat{\theta}_{0jt}$ and $\hat{\theta}_{1jt}$ are the final adaptive estimates at time t.

(2) DNS: the NS factors are obtained by fitting (2.5) to the data. The parameters of the stationary AR modeling for each NS factor are estimated recursively. At each time point t, all the available past observations are used for the estimation. We have the following $h-$step ahead forecasts:

$$\begin{aligned} \hat{X}_{t+h}(\tau) &= \hat{\beta}_{1,t+h} + \hat{\beta}_{2,t+h}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \hat{\beta}_{3,t+h}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right), \\ \text{with } \hat{\beta}_{j,t+h} &= \hat{\theta}_{0jt} + \hat{\theta}_{1jt}\beta_{j,t}, \end{aligned}$$

where $\hat{\theta}_{0jt}$ and $\hat{\theta}_{1jt}$ are now the estimates from AR modeling by using all the past information.

We compute 1-, 6- and 12-month ahead forecasts for each of the curves dated from Jan 1994 to Dec 2000 for the U.S. data and Jan 2008 to Oct 2011 for the China data respectively. Recursive forecasting with an extending window always starting from the first month is conducted. Take the U.S. data for an example. We use the yield curves from Jan 1985 to Dec 1993 to obtain the 1-month ahead forecast for Jan 1994, and the data from Jan 1985 to Jul 1993 to get the 6-month forecast for Jan 1994. We move forward one period at a time and redo the forecast until reaching the end of the sample. Therefore, in order to obtain the 1-month (6-month) ahead forecast for Feb 1994, data curves from Jan 1985 to Jan 1994 (Jan 1985 to Aug 1993) are utilized. In Figure 2.4.1 we depict the 1-, 6- and 12-month ahead out-of-sample forecasts for dates of July 1994 (U.S. Treasuries) and of April 2010 (China

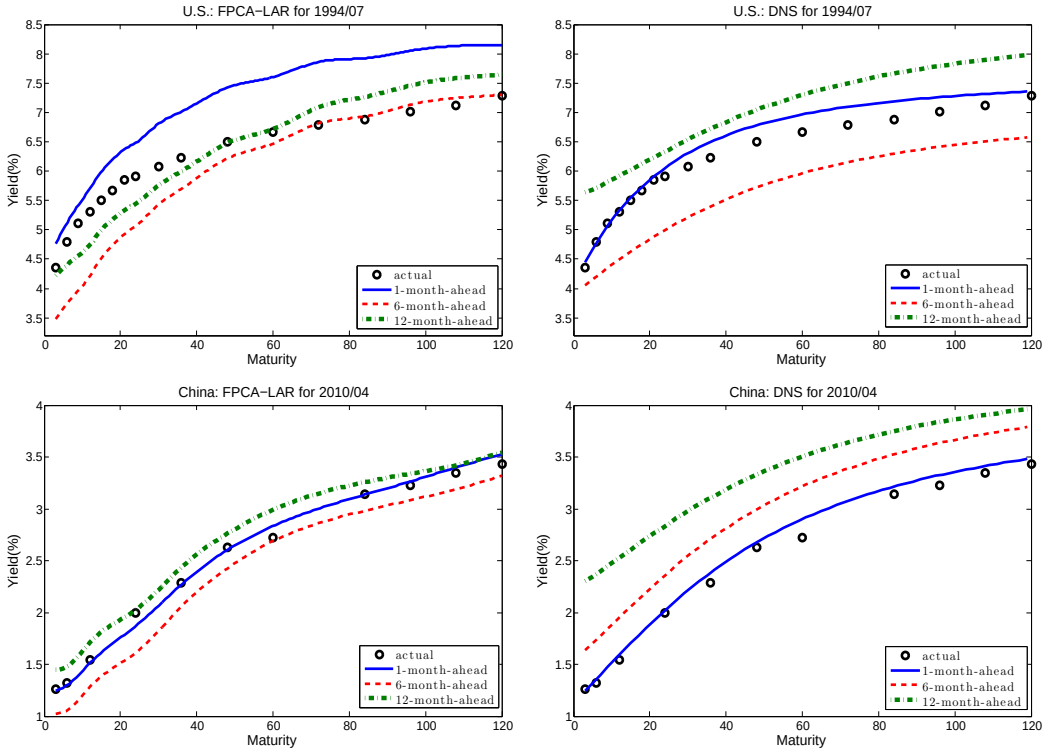

Figure 2.4.1 Out-of-sample forecasts. The actual discrete interest rates (dotted), the DNS forecast (right) and the FPCA-LAR forecast (left) for 1-, 6- and 12-month ahead horizons on dates July 1994 for U.S. market and April 2010 for China market.

Treasuries). On these two dates, the yield curves display typical shapes in the market, e.g., with a double-humped curvature shape in China market in April 2010. It illustrates that for both data the FPCA-LAR method performs well in 6- and 12-month ahead forecasts, whereas the DNS model provides accurate results in the immediate forecast horizon. To measure the accuracy, we compute the root mean squared forecast error (RMSFE) between the forecasts and the actual

values. Table 2.2 reports the average values at various maturities. The results also indicates that the proposed FPCA-LAR model is indeed superior to the DNS model in 6- and 12-month ahead forecasts, with smaller values of forecast errors.

The out-of-sample forecasts show that the proposed FPCA-LAR method attains overall better results than DNS model, which justifies that forecasting yield curves via the adaptive method is more flexible and accurate. For different kinds of yield curves, the FPCA factors would be a better choice especially for those markets whose underlying data generating processes deviate from exponential basis functions.

	U.S. yield curves						China yield curves					
h	1-month		6-month		12-month		1-month		6-month		12-month	
τ	DNS	F-L	DNS	F-L	DNS	F-L	DNS	F-L	DNS	F-L	DNS	F-L
3	0.17	0.24	0.56	**0.38**	0.87	**0.48**	0.29	0.31	0.99	**0.74**	1.11	**0.68**
6	0.20	0.23	0.61	**0.44**	0.86	**0.54**	0.28	0.30	0.99	**0.75**	1.11	**0.69**
9	0.22	0.24	0.64	**0.49**	0.86	**0.57**	-	-	-	-	-	-
12	0.24	0.27	0.67	**0.53**	0.87	**0.59**	0.28	0.28	0.95	**0.74**	1.06	**0.69**
15	0.25	0.30	0.69	**0.55**	0.88	**0.59**	-	-	-	-	-	-
18	0.26	0.31	0.71	**0.57**	0.90	**0.60**	-	-	-	-	-	-
21	0.27	0.33	0.73	**0.59**	0.93	**0.61**	-	-	-	-	-	-
24	0.28	0.34	0.76	**0.60**	0.97	**0.62**	0.28	**0.25**	0.89	**0.75**	0.96	**0.68**
30	0.28	0.38	0.76	**0.61**	1.00	**0.62**	-	-	-	-	-	-
36	0.28	0.40	0.77	**0.62**	1.03	**0.63**	0.25	**0.24**	0.78	**0.67**	0.81	**0.58**
48	0.28	0.47	0.77	**0.65**	1.08	**0.66**	0.23	0.25	0.67	**0.60**	0.69	**0.51**
60	0.29	0.52	0.81	**0.69**	1.16	**0.72**	0.25	0.28	0.63	**0.58**	0.63	**0.51**
72	0.27	0.56	0.78	**0.71**	1.16	**0.74**	-	-	-	-	-	-
84	0.27	0.56	0.77	**0.71**	1.17	**0.75**	0.23	0.29	0.57	**0.56**	0.57	**0.48**
96	0.26	0.59	0.74	**0.73**	1.15	**0.78**	0.21	0.29	0.54	0.55	0.53	**0.45**
108	0.25	0.61	0.74	0.75	1.17	**0.81**	0.19	0.28	0.50	0.52	0.51	**0.41**
120	0.26	0.62	0.76	**0.75**	1.22	**0.82**	0.19	0.30	0.46	0.49	0.47	**0.39**

Table 2.2 RMSFE: The average values of the out-of-sample forecast errors for the forecasting horizons h of 1-, 6- and 12-month ahead at various maturities, τ, from 3 months to 10 years. The DNS model and the FPCA-LAR (F-L) model are applied to U.S. Treasuries and China Treasuries. The better performance of F-L model is marked in bold.

CHAPTER 3

Multivariate model with LVAR

In this chapter, we propose a LVAR model to estimate the joint dynamics of multiple time series. The LVAR model allows parameters to be time dependent, without any particular assumptions on the variation type. Time-varying parameters at each point in time are, of course, too flexible to constitute an identifiable dynamic model. We therefore employ a local homogeneity assumption to balance between model flexibility and estimation feasibility. Local homogeneity assumes that at any particular time point there exists a past time interval, over which the local sample can be well approximated by a VAR model with constant parameters. A sequential testing procedure is used to find the longest interval that satisfies

the local homogeneity assumption and the interval of local homogeneity is thus identified.

Instead of using all the available past information, as in the recursive estimation approach, we conduct the estimation utilizing the interval of local homogeneity in the adaptive approach. The interval satisfying local homogeneity is time dependent and of possibly varying interval length. It is also different from the conventional rolling window or recursive window technique that adopts a fixed window size or expands the window size throughout the estimation. The fitted adaptive model is then used to monitor the model parameters and to forecast variables if the model is sufficiently parsimonious and the selected local interval is sufficiently long to include efficient information for accurate prediction. The proposed LVAR model is applied to U.S. Treasuries but at different dates and time to maturities compared to Chapter 2. Details of the method and applications will be discussed and demonstrated in later sections. The content and materials presented in this chapter have been published in Chen, Li and Niu (2013).

3.1. Method

3.1.1 Adaptive vector autoregressive model

Let $\boldsymbol{X}_t \in \mathbb{R}^d$ denote $d-$dimensional autoregressive time series variables, with $t = 1, \cdots, T$. The LVAR model is defined with time-varying parameters as:

$$\boldsymbol{X}_t = c_t + A_{1t}\boldsymbol{X}_{t-1} + \cdots + A_{pt}\mathbf{X}_{t-p} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_t),$$

where $c_t = (c_{1t}, \cdots, c_{dt})^\top$ is the intercept vector at time point t and A_{jt} is a $d \times d$ matrix for $j = 1, \cdots, p$. The stochastic innovation ϵ_t is assumed to be Gaussian distributed satisfying $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t \epsilon_t^\top) = \Sigma_t$. Moreover, we assume that there is no serial correlation between any two innovations across time, that is $\mathrm{E}(\epsilon_t \epsilon_{t-k}^\top) = 0$ for $k \neq 0$. The LVAR model, with time-varying parameters, is appropriate in a non-stationary situation where structural changes exist. It also works well in a homogeneous case by fixing the parameters as constant. In either case, the estimation is conducted under the local homogeneity assumption. That is, the multivariate time series are approximated by a parametric model over an interval of local homogeneity.

Although the order of the LVAR model, denoted by p, is allowed to be more than 1, we fix the lag order to 1. In addition, the adoption of the simplest model

structure is significantly motivated by tractability of monitoring and good out-of-sample forecast performance. In Section 3.2, we investigate the model misspecification issue in a simulation study, where the true data generating process has a higher order than the recommended one. It shows that the LVAR model of order 1 provides stable performance. The LVAR model of order 1, represented in matrix form, is as follows:

$$\boldsymbol{X}_t = c_t + A_t \boldsymbol{X}_{t-1} + \epsilon_t.$$

For notational simplicity, we denote the unknown parameters by $\Theta_t = (c_t, A_t, \Sigma_t)$.

3.1.2 Estimation under local homogeneity

Suppose that at time point t, the time series is homogeneous with $\Theta_t = \Theta$ over an interval $I_t = [t - m_t + 1, t]$. Parameter m_t is the interval length, corresponding to the number of observations in the local sample. The local (quasi-) log-likelihood function is defined as:

$$\ell\left(I_t, \Theta_t\right) = -\frac{m_t}{2} \ln |2\pi\Sigma_t| - \frac{1}{2} \sum_{s=t-m_t+1}^{t} \epsilon_s^\top \Sigma_t^{-1} \epsilon_s$$

from which we obtain the local maximum likelihood estimate (MLE):

$$\widetilde{\Theta}_t = \operatorname{argmax} \ell\left(I_t, \Theta_t\right).$$

Now we relax the local homogeneity assumption such that $\Theta_t \approx \Theta$. Then the

modeling bias of the parametric model with constant parameter Θ and the local parametric model with time dependent parameter Θ_t can be measured by:

$$\Delta_t = |\ell(I_t, \Theta) - \ell(I_t, \Theta_t)|^{1/2}$$

which should be small. Therefore, the local MLE, though not unbiased in this situation, can be used.

In practice, the interval of local homogeneity is unknown. The question is how to identify it or equivalently how to select interval length m_t at any particular time point t. With too large a value, there is a high probability of having non-trivial modeling bias, which violates the local homogeneity assumption. On the contrary, a small value of m_t though satisfying a small modeling bias, unnecessarily discards too many observations that are useful for estimation. The goal is to select the longest interval that does not violate the local homogeneity assumption.

Suppose there are K candidate intervals at time point t, which contain the interval of local homogeneity I_t:

$$I_t^{(1)} = [t - m_t^{(1)} + 1, t], \quad \cdots, \quad I_t^{(K)} = [t - m_t^{(K)} + 1, t]$$

with $I_t^{(1)} \subset \cdots \subset I_t^{(K)}$. A sequential testing procedure helps to select the longest interval that satisfies local homogeneity. It is worth noting that beyond the selected interval, there is a high probability of structural changes. The adaptive technique proceeds as follows. The procedure starts from the shortest interval $I_t^{(1)}$, over which

the local homogeneity assumption probably holds. We accept the interval and denote the accepted adaptive estimate by $\hat{\Theta}_t^{(1)} = \widetilde{\Theta}_t^{(1)}$. Iteratively, the procedure extends to the next interval $I_t^{(k)}$ for $k \geq 2$, where we introduce the test statistic:

$$T_t^{(k)} = \left| \ell\left(I_t^{(k)}, \widetilde{\Theta}_t^{(k)}\right) - \ell\left(I_t^{(k)}, \hat{\Theta}_t^{(k-1)}\right) \right|^{1/2},$$

which measures the divergence of the hypothetical model from the recently accepted local model. If the divergence is significant, where the significance level is controlled by a critical value ζ_k, it indicates that there is a significant structure change larger than the one arising due to sampling changes. In this case, we reject the null hypothesis of local homogeneity and terminate the selection procedure. The last accepted interval would be the final selection and we have $\hat{\Theta}_t^{(j)} = \hat{\Theta}_t^{(k-1)}$ for $j = k, \cdots, K$. Here j denotes the index of the candidate interval where the procedure is terminated. Otherwise, we accept the longer interval $I_t^{(k)}$, and update the adaptive estimate by $\hat{\Theta}_t^{(k)} = \widetilde{\Theta}_t^{(k)}$. The test procedure is continued on the next interval until either a change is detected or the longest candidate interval is reached.

3.1.3 Calibrate critical values

The success of the adaptive selection procedure depends on the critical values, which are calibrated in Monte Carlo experiments. As the critical values control

the significance level under the local homogeneity assumption that requests a small modeling bias, we generate samples with homogeneity and measure the modeling bias using the adaptive estimation. The critical values are selected such that they are capable of providing the prescribed performance of the testing procedure.

The homogeneous VAR processes are generated with constant parameters $\Theta^* = (c^*, A^*, \Sigma^*)$, such that

$$\boldsymbol{X}_t = c^* + A^* \boldsymbol{X}_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma^*).$$

Each process includes T observations and the generation is repeated N times. For each generated process $\mathbf{X}_{1:T}^n$, $n = 1, ..., N$, the same interval set $\{I_t^{(k)}\}_{k=1}^K$ is used everywhere for $t = m^K, \cdots, T$ and $k = 1, \cdots, K$, where m^K is the longest interval length for interval $I_t^{(K)}$. In the following, for ease of elaboration, we drop the series index n. Under the assumption of homogeneity, the estimation error can be measured by the fitted log-likelihood ratio over each interval:

$$R_k = E_{\Theta^*} \left| \ell \left(I^{(k)}, \tilde{\Theta}_t^{(k)} \right) - \ell \left(I^{(k)}, \Theta^* \right) \right|^{1/2}, \tag{3.1}$$

where R_k can be computed numerically with knowledge of Θ^*.

Once a set of critical values $\zeta_1, \cdots, \zeta_K$ is given, one can employ the adaptive procedure, by checking the significance of the test statistic $T_t^{(k)}$, to obtain the adaptive estimate $\hat{\Theta}_t^{(k)}$ of the time-dependent parameter Θ_t. Given the MLE $\tilde{\Theta}_t^{(k)}$ of the constant parameter Θ^*, the temporal realized modeling bias can be measured

as:

$$\delta_t^{(k)} = \left| \ell\left(I^{(k)}, \tilde{\Theta}_t^{(k)}\right) - \ell\left(I^{(k)}, \hat{\Theta}_t^{(k)}\right) \right|^{1/2}.$$

The adaptive estimation should behave as well as the true underlying characteristics under the null of time homogeneity, in the sense that the modeling bias is bounded by the estimation error R_k in Equation (3.1) with knowledge of the true model. Therefore, we have

$$E_{\Theta^*}\left(\delta_t^{(k)}\right) = E_{\Theta^*}\left| \ell\left(I^{(k)}, \tilde{\Theta}_t^{(k)}\right) - \ell\left(I^{(k)}, \hat{\Theta}_t^{(k)}\right) \right|^{1/2} \leq R_k. \tag{3.2}$$

Clearly, the critical values are the only unknown parameters in the above inequality Equation (3.2), which can be calibrated.

The computation of critical values relies on two hyperparameters, the interval candidates $\left(I_t^{(1)}, \cdots, I_t^{(K)}\right)$ and Θ^*. In our study, at each point of time, we consider $K = 19$ intervals for the adaptive estimation, starting with 12 months and a continuous increment of $M = 6$ months between any adjacent intervals, i.e., 120 months (10 years) is the maximal sample size. Ideally, Θ^* should be close to the true parameter underlying the real data series at each point of time, which is actually the target of our estimation. In practice, we approximate Θ^* with the estimate from a sub-sample, for example, the sub-sample before the forecast exercise starts. We find that the adaptive technique is quite robust to the selection of the hyperparameters, as is illustrated in Section 3.2. There is no significant difference in terms of forecast accuracy for different sets of interval candidates determined by K and M as well as

for possible misspecifications of Θ^* with $\pm 20\%$ deviation from the true values. In the following, we use the MLE of the available real sample of U.S. Treasuries from January 1983 to December 1997 as Θ^*, i.e., using all the information before the forecast exercise starts. We then generate homogeneous VAR series, and calibrate the set of critical values as described above. The same set of calibrated critical values is adopted for every time point throughout the real-time estimation and forecast.

The adaptive estimation algorithm for any particular time point t is as follows:

(1) Calibrate critical values $\{\zeta_k\}_{k=1}^K$ given a set of interval candidates $\left\{I_t^{(k)}\right\}_{k=1}^K$.

- Generate homogeneous VAR processes with constant parameter Θ^*. We use the MLE for data sample before the first forecast origin.
- Compute MLEs $\tilde{\Theta}_t^{(k)}$ and the risk bound R_k over each interval candidate.
- Given an initial set of critical values, obtain the adaptive estimator $\hat{\Theta}_t^{(k)}$. Compute the realized modeling bias $\delta_t^{(k)}$ and check the risk bound as in (3.2). If it holds, reduce the critical values. Otherwise, increase the critical values. Repeat until the cutting point is found.

(2) Given the calibrated critical values, for the data set of interest to be investigated, conduct the sequential testing procedure to identify the interval of

local homogeneity and estimate the adaptive estimator over time. Starting from an initial time t_0, for $t \geq t_0$:

- **Initialization:** We accept the shortest interval and set $\hat{\Theta}_t^{(1)} = \widetilde{\Theta}_t^{(1)}$.
- **Loop:** For $k \geq 2$,

  if $T_t^{(k)} = \left| \ell\left(I_t^{(k)}, \widetilde{\Theta}_t^{(k)}\right) - \ell\left(I_t^{(k)}, \hat{\Theta}_t^{(k-1)}\right) \right|^{1/2} \leq \zeta_k$, we accept the interval $I_t^{(k)}$ and update the estimate:

  $$\hat{\Theta}_t^{(k)} = \widetilde{\Theta}_t^{(k)}.$$

  Otherwise, terminate the procedure and we have:

  $$\hat{\Theta}_t^{(j)} = \hat{\Theta}_t^{(k-1)}, \quad j = k, \cdots, K.$$

- **Final estimate:** $\hat{\Theta}_t = \hat{\Theta}_t^{(K)}$.

(3) We assume that the interval of local homogeneity is extendable over the forecasting horizon, denoted by h. The fitted LVAR model is used for the prediction: $\hat{\boldsymbol{X}}_{t+h} = \hat{c}_t + \hat{A}_t \boldsymbol{X}_t$.

3.2. Simulation

In this section, we conduct a simulation study to demonstrate the performance of the LVAR model. In particular, we evaluate the forecast accuracy of the adaptive procedure, compared to alternative methods with window length selection

such as the rolling window technique. Furthermore, the robustness of the forecast performance for the adaptive procedure is investigated with respect to the choice of hyperparameters (Θ^*, K, M). Moreover, we address the model misspecification issue of whether the simplest adaptive model of order 1 is sufficient to handle autoregressive processes of a higher order.

3.2.1 Simulation design

We consider two kinds of scenarios: a homogeneous scenario with globally constant parameters, denoted as HOM; and a heterogeneous scenario with time-varying parameters shifting from one level to another, i.e., a regime switching scenario denoted as RS.

In the HOM scenario, we calibrate the VAR coefficients from a three-factor VAR(1) model constructed by Nelson-Siegel yield factors with U.S. yield curve data from 1983 to 1997. The NS model is parameterized according to Diebold and Li (2006), and the data set contains fifteen yield series as used in Chen and Niu (2014). We denote the underlying parameters as $\Theta_0 = (c_0, A_0, \Sigma_0)$ and keep them constant throughout the whole sample. In the RS scenario, we design two experiments and label them as: RS-A where A denotes the VAR coefficient matrix, and RS-C where C denotes the intercept vector. Only the labeled parameters shift

Scenarios	Parameters	
HOM	$c_0 = \begin{pmatrix} 0.093 \\ 0.111 \\ -0.314 \end{pmatrix}$ $A_0 = \begin{pmatrix} 0.989 & 0.011 & -0.005 \\ -0.031 & 0.933 & 0.054 \\ 0.062 & 0.090 & 0.853 \end{pmatrix}$	
RS	Phase 1 $t \in [1, 200]$	Phase 2 $t \in [201, 400]$
RS-A	$A_t = \begin{pmatrix} 0.989 & 0.011 & -0.005 \\ -0.031 & 0.933 & 0.054 \\ 0.062 & 0.090 & 0.853 \end{pmatrix}$	$A_t = \begin{pmatrix} 0.493 & -0.167 & 0.177 \\ 0.259 & 0.952 & -0.082 \\ 0.523 & 0.511 & 0.462 \end{pmatrix}$
RS-C	$c_t = (0.093, 0.111, -0.314)^\top$	$c_t = (2.789, -1.974, -3.503)^\top$

Table 3.1 Parameters in the simulation scenarios. HOM refers to the homogeneous scenario; and RS refers to the regime-switching (structural change) scenario. In each of the RS scenarios, only the labeled parameter is changed in Phase 2. The other parameters remain the same as in the original set-up.

from the original level to a new set of parameter values estimated using a different subsample of the NS factors, during the recent financial crisis from 2008 to 2010. The other parameters remain the same as in the original set up. For each scenario and experiment we simulate 200 data series, each with 400 observations. In the RS scenarios, each of the regimes lasts for 200 time points. The details of the scenario designs regarding timing and parameters are described in Table 3.1.

We employ both the adaptive procedure and the rolling window strategies to compute the one-step ahead forecasts for the same forecasting period from $t = 122$ to 400 for each of the simulated samples. In the adaptive case, the critical values are calibrated using the true underlying parameters, i.e., $\Theta^* = \Theta_0$. The candidate intervals start from 12 months (1 year) and end at 120 months (10 years), with $K = 19$ and $M = 6$. The interval lengths are 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90, 96, 102, 108, 114, 120. Figure 3.2.1 displays the resulting critical values that are used in each scenario. At the same time, we consider 19 alternative

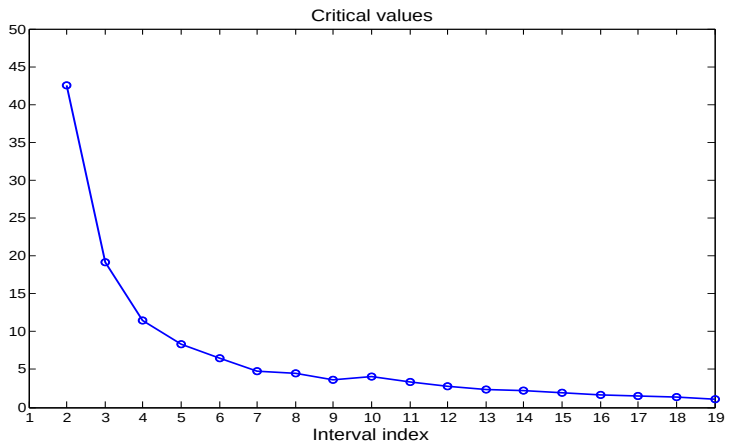

Figure 3.2.1 Critical values. The hyperparameters are $M = 6$, $K = 19$ and $\Theta^* = \Theta_0$.

window sizes in the prediction using the rolling window technique, i.e., ranging from I^1 (12 months) to I^K (10 years), which correspond to our interval candidates in the adaptive procedure. Forecast accuracy is determined by the forecast root

mean squared error (RMSE) values.

3.2.2 Forecast accuracy

Table 3.2 presents the RMSE values of the adaptive technique and the rolling window alternatives for all the three NS factors NS1, NS2 and NS3. For ease of exposition we do not report all forecasting results of the rolling windows. Instead, we only list those window sizes yielding the best forecast accuracy (with minimal RMSE values) and the worst accuracy (with maximum RMSE values). The respective rolling window sizes are indicated in the parentheses. The number of times that the LVAR is superior to the 19 alternative window choices is highlighted in the column of "No. of Winning".

The numerical results reveal that the adaptive approach with varying local window sizes introduces more flexibility into the procedure, leading to a comparable performance to the optimal sample under the homogeneous scenario and a generally better performance under the scenarios with structural changes. More specifically, in the homogeneous scenario, the adaptive technique, though with a misspecified assumption of time-varying coefficients, still provides reasonable accuracy. In the structural change scenarios with time-varying parameters, our technique is superior to almost all the 19 alternative rolling window estimations.

	Scenario	Rolling window (window size) Best	Worst	Adaptive	No. of Winning
NS1	HOM	0.336 (120)	0.424 (12)	0.340	13/19
	RS-A	0.400 (78)	0.517 (12)	0.395	19/19
	RS-C	0.501 (54)	0.606 (12)	0.489	19/19
NS2	HOM	0.370 (120)	0.474 (12)	0.374	14/19
	RS-A	0.399 (120)	0.514 (12)	0.398	19/19
	RS-C	0.472 (72)	0.588 (12)	0.465	19/19
NS3	HOM	0.816 (120)	1.049 (12)	0.826	13/19
	RS-A	0.851 (120)	1.084 (12)	0.855	16/19
	RS-C	0.953 (96)	1.195 (12)	0.954	16/19

Table 3.2 Forecast accuracy. The rolling window adopts one of the predetermined window lengths of $k \times M$, where $k = 1, \cdots, 19$ and $M = 6$, throughout the whole sample. The adaptive technique adopts a selected time-varying window length among the choices of the interval sets at each point of time. For the performance of the rolling windows, only the best and worst results with the related window choices are reported. We also report the number of wins of the adaptive technique compared to the 19 rolling window estimation alternatives.

The adaptive interval selection procedure contributes to the improvement of forecast accuracy and simultaneously provides stable performance. In the structural change scenarios with a parameter shift at $t = 201$, the average values of the selected intervals drop quickly after that point, see Figure 4.3.2. As the sample

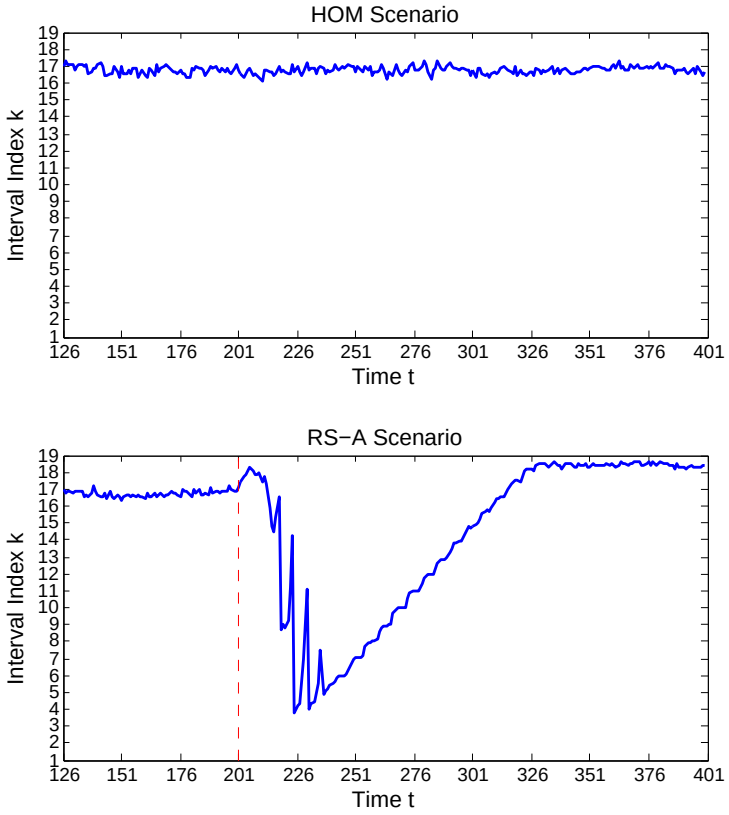

Figure 3.2.2 The average values of the selected intervals from time index 122 to 400 over the 200 generated processes in the HOM and RS-A scenarios.

following the new data generating process extends, the lengths of the selected intervals increase. The conventional rolling window technique with a fixed window size, on the other hand, does not have such flexibility. Moreover, in the homogeneous scenario where there is no structural changes, the optimal interval selection should be the longest one of 120 months. The average values of the selected intervals are quite reasonable, with values around 108 months ($k = 17$) for each time point. A

direct comparison confirms that the simple yet flexible LVAR model can be safely applied to both stationary and non-stationary situations.

3.2.3 Robustness check

The forecast accuracy of the adaptive model depends on the critical values which themselves depend on the underlying hyperparameters K, M and Θ^*. As an illustration, we analyze the robustness of the hyperparameter choices under the RS-A scenario. The default values are $K = 19$, $M = 6$ and $\Theta^* = \Theta_0$. We report the forecast accuracy with different hyperparameters and also analyze the impact on the forecast performance when Θ^* is misspecified.

We first consider four alternative interval sets: given $M = 6$, taking fewer or more candidates with $K = 10$ or $K = 30$; and given $K = 19$ with the first interval being 12 months, taking shorter or longer steps with $M = 3$ or $M = 12$ between two adjacent intervals. With these alternative interval sets, the critical values are re-calibrated. In order to match the longest possible interval length for $I^{(K)}$, the initial forecasting points for cases with $K = 10$ and $K = 30$ are different, which are $t = 122$ and $t = 188$, respectively.

Moreover, we consider the cases where parameter Θ^* is misspecified. Instead of using the true underlying parameter values, i.e., $\Theta^* = \Theta_0$, we compute critical

values under each of the two misspecified hypothetical parameters, i.e., ±20% deviation from the true VAR coefficients. More specifically, we decompose matrix A and shift its eigenvalues by 20% to have $0.8 \times$ EV and $1.2 \times$ EV, and denote the scenarios as mis08 and mis12, respectively. We use these parameter sets to generate Monte Carlo experiments and to calibrate the critical values, although the series actually follow the VAR model with Θ_0. A potential problem occurs in scenario mis12, where stationarity is not valid even in the shortest interval. To guarantee the existence of local homogeneity, we artificially select a matrix that satisfies the stationarity condition. The forecast accuracy of scenario RS-A with the new hyperparameters is computed and compared again with the alternative rolling window forecasts.

Table 3.3 presents the forecast accuracy under the alternative or misspecified hyperparameters. The results confirm the robustness of the adaptive technique, with RMSE values very close to those in the default case where true parameter values are used to calibrate the critical values.

3.2.4 Model misspecification

In the following experiment, we investigate the stability of the proposed adaptive model in terms of model misspecification. It is necessary to answer whether

	default	$K = 10$	$K = 30^{*1}$	$M = 3$	$M = 12$	default^{*2}	mis08^{*2}	mis12^{*2}
NS1	0.395	0.395	0.409	0.396	0.397	0.348	0.349	0.347
NS2	0.398	0.402	0.403	0.400	0.399	0.378	0.381	0.376
NS3	0.855	0.862	0.862	0.858	0.857	0.844	0.849	0.841

Table 3.3 Robustness testing (scenario RS-A): RMSE values. We compare the default case of $M = 6$, $K = 19$ and $\Theta^* = \Theta_0$ to several cases of alternative hyperparameters of $M = 3$ or 12, $K = 10$ or 30 and misspecified parameter Θ^* in the critical value calibration.

^{*1}The first forecast is at time index 188 (instead of 122 as for others) in order to correspond to the longest possible interval length.

^{*2}An artificial VAR coefficient matrix is used to guarantee the existence of local homogeneity after being multiplied by 120% in the mis12 scenario.

the adaptive model of order 1 provides reasonable forecast accuracy, if the true data generating process has a higher lag order. As an illustration, we consider the true data generating process to be a LVAR process with lag order 5. Among other scenarios, we conduct the simulation under the RS-A scenario where the VAR coefficient matrix has regime shifts. In the first regime, the parameter set is computed from a VAR(5) model using the three NS factors extracted from the U.S. yield curve from 1983 to 1997. The autoregressive coefficient matrix changes to a new value estimated by using a different sample from 1983 to 2010. Prediction

exercises are conducted using both a LVAR(5) model and a misspecified model, the LVAR of order 1. In order to guarantee a reasonable estimation result for VAR modeling of order 5 over the first few intervals, we select the third one from the default interval candidates as the first interval in this section. In other words, we have now 17 interval candidates with the shortest length being 24 months and the longest one being 10 years. The increment between adjacent intervals remains to be 6 months.

Table 3.4 displays the forecast performance based on both the correct and the misspecified models of LVAR(5) and LVAR(1), respectively. The adaptive models are in most cases superior to the alternatives using various rolling window approaches. Misspecified model LVAR(1) provides even better accuracy with smaller RMSE values than LVAR(5). This implies that the proposed LVAR(1) model is capable of providing reasonable forecast accuracy even when it is misspecified. The simple structure is beneficial for out-of-sample forecasts.

3.3. Real data analysis

In this section, the proposed LVAR model is fitted to monitor and forecast the U.S. Treasuries, spanning from January 1983 to September 2010. In each month, there is a yield curve for interest rates at 15 maturities (3, 6, 9, 12, 18, 24, 30, 36,

	Scenario	Rolling window (window size) Best	Worst	Adaptive	No. of Winning
NS1	$p = 1$	0.385 (60)	0.406 (24)	0.378	17/17
	$p = 5$	0.403 (120)	0.706 (24)	0.409	11/17
NS2	$p = 1$	0.388 (84)	0.412 (24)	0.383	17/17
	$p = 5$	0.408 (120)	0.695 (24)	0.416	11/17
NS3	$p = 1$	0.822 (84)	0.878 (24)	0.814	17/17
	$p = 5$	0.862 (120)	1.458 (24)	0.876	11/17

Table 3.4 Model misspecification with the true data generating process of LVAR(5). In the table, $p = 1$ and $p = 5$ refer to the misspecified and correct lag orders, respectively. Only the best and worst results of all the rolling window approaches (with the corresponding window sizes) are reported. The last two columns contain the LVAR results and the number of cases where LVAR is better than the rolling window approaches in terms of RMSE values.

48, 60, 72, 84, 96, 108 and 120 months). The short-term yields at 3 and 6 months are converted from the 3- and 6-month Treasury Bill rates on a discount basis, available from the Federal Reserve's H.15 release of selected interest rates. The remaining yields with maturities of integer years are taken from publicly available research data of the Federal Reserve Board, as released by Gürkaynak, Sack and Wright (2007). We add the 9-, 18-, and 30-month yields interpolated according to the parameters provided in their data file to emphasize the fit for medium-term yields. The time evolution of the yield curves is displayed in Figure 3.3.1

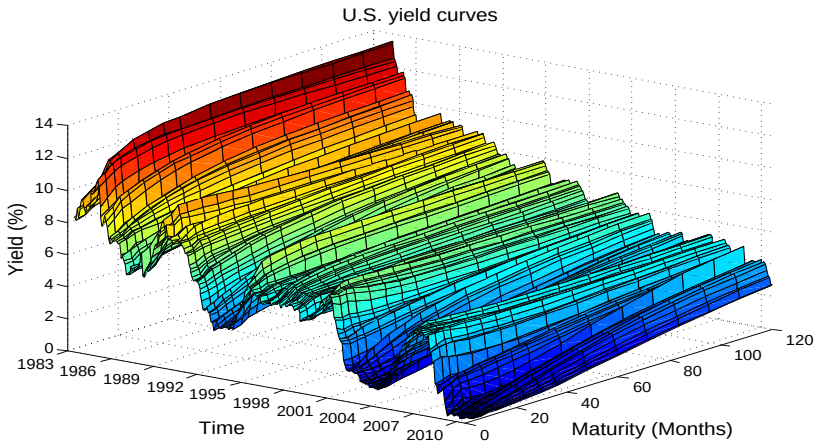

Figure 3.3.1 Time evolution of the U.S. yield curves from January 1983 to September 2010.

The three NS factors are extracted in the framework of Nelson and Siegel (1987):

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \epsilon_t(\tau), \qquad (3.3)$$

where $\epsilon_t(\tau) \sim N(0, \sigma_\epsilon^2)$ and $y_t(\tau)$ denotes the yield curve with maturity τ (in months) at time t. We follow Diebold and Li (2006) to set $\lambda = 0.0609$ which maximizes the curvature loading at a medium maturity of 30 months. The three factors β_{1t}, β_{2t} and β_{3t}, represent level, slope and curvature, respectively. The dynamics of the three factors are displayed in Figure 3.3.2.

Under the NS framework with a fixed value of λ, if there exists any form of non-stationarity in the yield curves y_t, then it is solely attributed to changes in the sequences of the state factors, denoted as $\mathbf{X}_t = [\beta_{1t}, \beta_{2t}, \beta_{3t}]'$. The sample

autocorrelation and cross-correlation plots of the factors are presented in Figure 3.3.3. It shows that the NS factors are cross-dependent not only on its own lagged values but also on the other factors. Moreover, these factors are persistent, with slowly decaying and significant autocorrelations and cross-correlations up to high lag orders, which cannot be easily captured by a VAR with a low order. The persistence feature motivates us to employ the LVAR model to describe and forecast the factors.

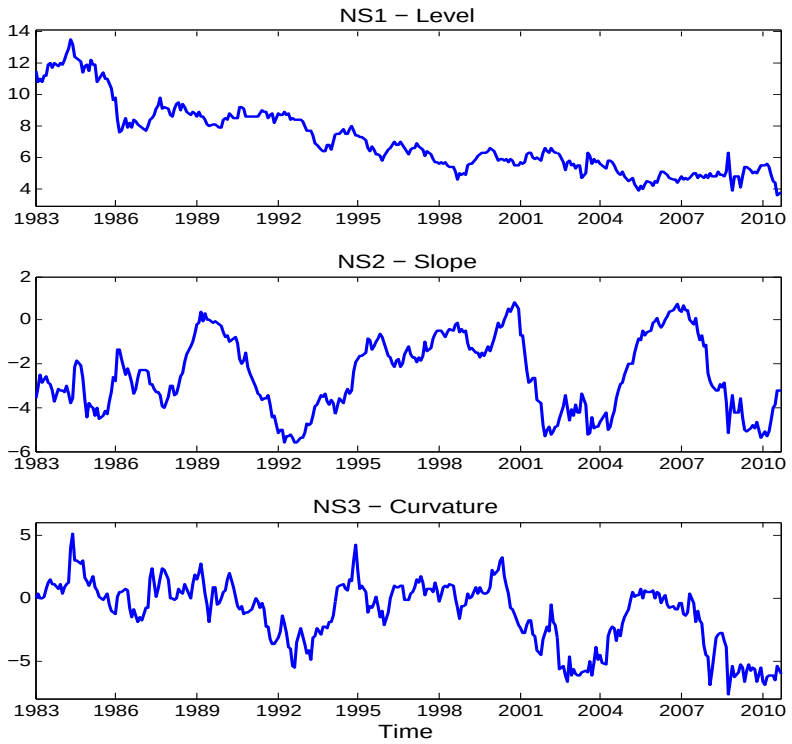

Figure 3.3.2 Time evolution of Nelson-Siegel factors extracted from U.S. yield curves.

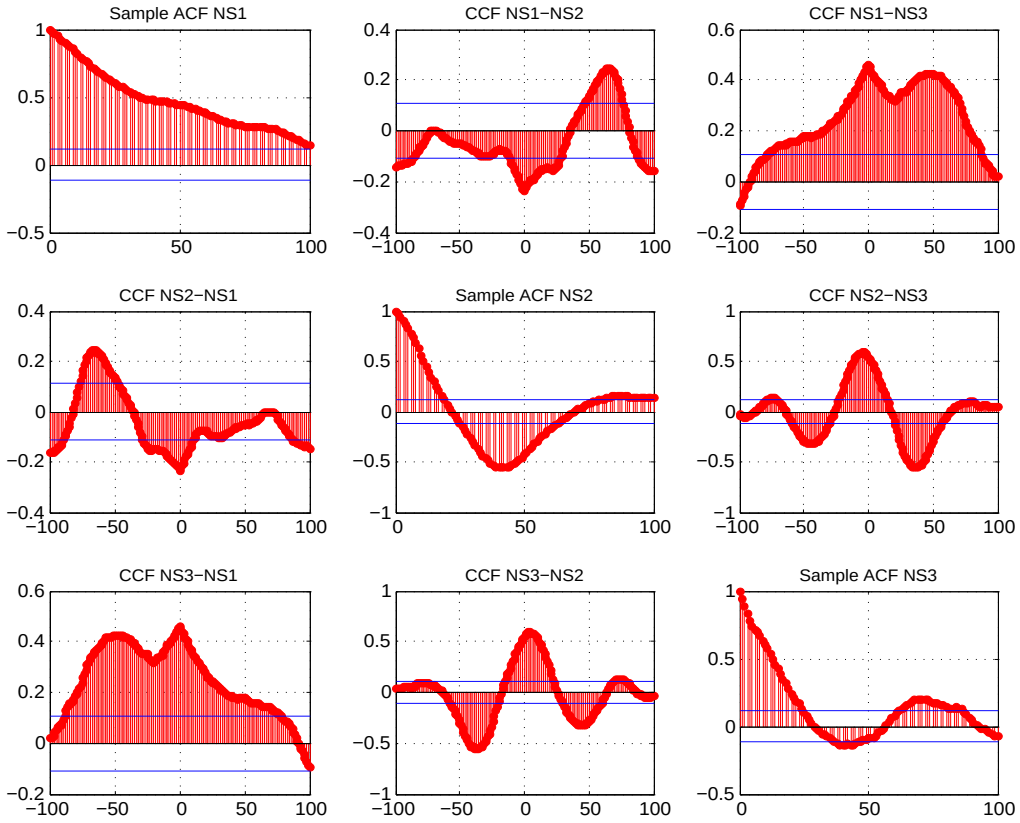

Figure 3.3.3 Sample autocorrelations and cross-correlations of the three NS factors.

At any forecast time, we use the adaptive technique to identify an interval of local homogeneity, over which we estimate the parameters. The fitted model is then used to iteratively compute multi-step ahead forecasts of the NS factors, which are further used to obtain the forecasts of interest rates at various maturities in the NS framework. The first forecast originates from December 1997, where we iteratively obtain the 1-month to 12-month ahead forecasts for January 1998 to

December 1998. We then move to the next period to identify the interval of local homogeneity and fit the LVAR model for another set of forecasts up to a 12-month ahead forecast. This estimation and interval selection exercise are executed until September 2009 for the last set of forecasts between October 2009 and September 2010. A total of 142 forecasts are obtained for each forecast horizon.

The identified intervals of local homogeneity for dates from December 1997 (the first forecast origin) to September 2009 are shown in Figure 3.3.4. The plot shows that as the estimation moves forward, there are some commonly identified ending periods of homogeneous intervals, such as 1990, 1993 and 2000–2001, of which the timing of 1990 and 2001 coincide with U.S. economic recessions. In addition, the forecast results of LVAR are compared with the alternative forecasts using the rolling window technique and the recursive approach. The rolling window technique adopts a fixed window size, 60-month and 120-month in our study, while the recursive approach uses all the past available information. Table 3.5 summarizes the forecast performance of the three models for NS factors NS1, NS2 and NS3. Table 3.6 shows the multi-step ahead forecast results of the U.S. yields at 3-month, 12-month, 36-month, 60-month and 120-month maturities. Similar interpretations could be drawn from the two forecast results tables. In several cases, LVAR performs better than the rolling window technique with fixed window size of 60 months, showing smaller RMSE values. However, generally speaking, further

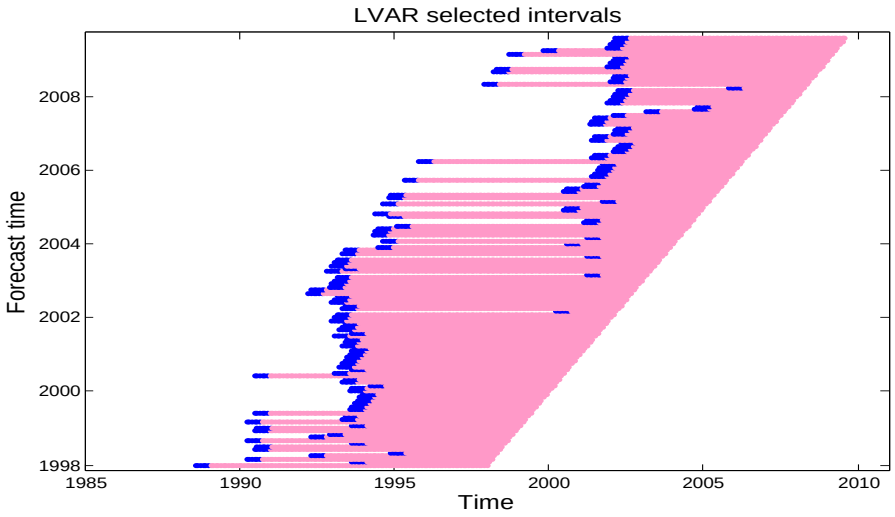

Figure 3.3.4 Selected intervals of local homogeneity for dates from December 1997 to September 2009. Over the intervals, the parameters are estimated and the fitted model is used to obtain the iterative forecasts. The vertical axis represents the time when the estimation and forecast are made. The selected interval is marked horizontally as a light pink line. The dark blue line represents the interval during which the most recent break is detected.

improvement of the LVAR model is needed to beat the rolling window model with 120-month maturity and the recursive approach.

In fact, as a specific case of this general setting, Chen and Niu (2014) show that restricting the state dynamics to an AR(1) model for each NS factor greatly improves forecast accuracy, and that the resulting performance beats the alternative rolling or recursive forecast uniformly. The reason may be due to the off-diagonal

elements in the VAR coefficient matrix being typically sparse and close to zeros, which does not contribute much to the improvement of forecast accuracy, but deteriorates the information efficiency. However, as a general illustration of LVAR application, we believe that this example provides a good scenario for monitoring and forecasting economic and financial time series.

NS1	h=1	h=3	h=6	h=12
LVAR	0.337	0.487	0.680	0.932
VAR Rolling 60m	0.337	0.475	0.636	0.783
VAR Rolling 120m	0.332	0.461	0.641	0.884
VAR Recursive	0.331	0.437	0.573	0.731
NS2	h=1	h=3	h=6	h=12
LVAR	0.416	0.717	1.178	2.294
VAR Rolling 60m	0.427	0.742	1.219	2.257
VAR Rolling 120m	0.417	0.687	1.069	1.844
VAR Recursive	0.412	0.662	1.003	1.747
NS3	h=1	h=3	h=6	h=12
LVAR	0.976	1.720	2.476	3.832
VAR Rolling 60m	0.927	1.635	2.375	3.288
VAR Rolling 120m	0.886	1.511	2.160	2.864
VAR Recursive	0.875	1.453	2.013	2.638

Table 3.5 RMSE values of the iterative forecasts for NS factors NS1, NS2 and NS3. Three types of models are employed: the LVAR model with a time-dependent interval of local homogeneity, a VAR rolling model with window sizes of 60 months and 120 months, and a recursive VAR model.

y(3)	h=1	h=3	h=6	h=12
LVAR	0.285	0.588	1.033	2.061
VAR Rolling 60m	0.281	0.586	1.044	1.896
VAR Rolling 120m	0.273	0.531	0.900	1.613
VAR Recursive	0.268	0.512	0.858	1.545
y(12)	h=1	h=3	h=6	h=12
LVAR	0.286	0.660	1.153	2.147
VAR Rolling 60m	0.276	0.631	1.103	1.882
VAR Rolling 120m	0.261	0.578	0.987	1.665
VAR Recursive	0.262	0.562	0.930	1.559
y(36)	h=1	h=3	h=6	h=12
LVAR	0.344	0.707	1.115	1.861
VAR Rolling 60m	0.326	0.664	1.031	1.557
VAR Rolling 120m	0.322	0.644	0.998	1.509
VAR Recursive	0.318	0.615	0.916	1.351
y(60)	h=1	h=3	h=6	h=12
LVAR	0.350	0.650	0.977	1.510
VAR Rolling 60m	0.336	0.609	0.888	1.240
VAR Rolling 120m	0.333	0.605	0.901	1.302
VAR Recursive	0.326	0.570	0.814	1.132
y(120)	h=1	h=3	h=6	h=12
LVAR	0.302	0.486	0.706	0.994
VAR Rolling 60m	0.299	0.460	0.626	0.769
VAR Rolling 120m	0.297	0.455	0.652	0.910
VAR Recursive	0.304	0.446	0.601	0.778

Table 3.6 RMSE values of the iterative forecasts for yields at 3-month, 12-month, 36-month, 60-month and 120-month maturities. Three types of models are employed: the LVAR model with a time-dependent interval of local homogeneity, the VAR rolling model with window sizes of 60 months and 120 months, and the recursive VAR model.

CHAPTER 4

Functional model with AFAR

The proposed AFAR modeling aims to model and forecast data that exhibit a high degree of cross-sectional dependence under the existence of non-stationarity. For the serial dependence, we take the California electricity prices as an example. Figure 1.3.1 displays the log-prices of the California electricity market that are observed 24 hours a day, 7 days a week from 5 July 1999 to 11 June 2000. The autocorrelations and cross-correlations among each hourly log-price are found to be significant. As an illustration, the sample autocorrelations of the hourly prices at 9am and the sample cross-correlations between the log-prices at 8am and 9am have been displayed in Figure 1.2.1, which implies strong serial dependence of the

two series. VAR modeling can be used to model the 24 multiple series, see Lütkepohl (2005); however, this kind of model is subject to the curse of dimensionality and hence leads to a lack of satisfactory accuracy in estimation and prediction. Alternatively, if we consider the multiple series of hourly electricity prices as functional observations that are defined in an infinite space, FDA may be used. On each day, there is one realization of daily electricity price curve smoothed over 24 observations (see smoothed log-price curves in Figure 1.3.1). With this natural and parsimonious representation of high-dimensional data, estimation and forecast accuracy may be improved. The FAR model, as an extension of VAR from a finite space to an infinite space, has been documented in literature, see e.g., Bosq (2000) for an extensive theoretical study.

Under the assumption that data are generated by a stationary FAR process with a constant mean function and serial dependence, the estimation of FAR operator is of great research interest. The constant mean of functional data, on the other hand, is less important under stationarity and ignored in the existing studies. For the existing methods on estimating the FAR operator, we refer to Section 1.3 for the literature review. It is interesting to note that maximum likelihood (ML) estimation is not privileged in estimating FAR model, possibly because the likelihood function is not well-defined in infinite dimensional parameter spaces. Nevertheless, based on Grenander (1981) theory of sieves, Geman and Hwang

(1982) propose a way to approximate the likelihood on the parameters' subspaces or sieves and establish generally consistent results. In the recent Mourid and Bensmain (2006), an ML estimator of the FAR operator is developed where the likelihood function is defined on sieves. However, the stochastic variations cannot be estimated given the definition of the likelihood function in their study; and they set the mean function to zero, which is inappropriate for non-stationary data.

Most existing works assume stationarity and have been implemented in different areas, e.g., analyzing the El Niño sea surface temperature index (see Besse et al., 2000; Antoniadis and Sapatinas, 2003; Mas and Pumo, 2007), as well as studying the term structure of Eurodollar futures rates (Kargin and Onatski, 2008). However, unstable turbulence does occur in practice, especially in economic and finance fields. The California electricity market, for example, is famous for the market crash of 2000, which leads to blackouts in the San Francisco area in January 2001 and the first bankruptcy in history of a power exchange, see Weron (2007). If functional data are not stationary, the proposed YW and ML estimators are biased with a high probability. To the best of our knowledge, the non-stationary issues haven't been well studied in functional data analysis, though we note that Besse et al. (2000) address the heteroscedasticity of covariance operator using a location-dependent weighting scheme and Horváth et al. (2010) propose a stability test of the FAR operator against a change point alternative. On the other hand,

methodologies for univariate non-stationary time series have been well developed. Among others, an adaptive approach with simple underlying models has attracted much attention, which is appropriate for various sources of changes without explicitly assuming change types and timing, see Belomestny and Spokoiny (2007) and Chen et al. (2010).

In the following, we will introduce the adaptive functional model, that is, the AFAR modeling. In Section 4.1, we derive the sieve estimators of the FAR operator, mean function and stochastic variations. Asymptotic consistency results are developed and proved. The adaptive estimation procedure and the associated theoretical properties are discussed in Section 4.2. In Section 4.3 we demonstrate the finite performance of the estimators in both stationary and non-stationary situations. Section 4.4 provides the real data analysis, where we implement the AFAR modeling to forecast the Californian electricity price curves. The forecast accuracy is compared with several alternatives including the univariate, multivariate and functional models.

4.1. FAR modeling under stationarity

We will first define the stationary FAR model and discuss about the method of sieves used in the ML estimation of the FAR operator, which is constant. After the

estimation and consistency results are established, the assumption of stationarity will be relaxed and the proposed adaptive modeling with time-varying operator ρ_t will be introduced. Details of the adaptive modeling AFAR can be referred to in Section 4.2.

As mentioned in Section 1.3, $\{X_1, ..., X_n\}$ denotes a sample of random curves with sample size n. Let $(\mathscr{H}, B_{\mathscr{H}})$ be a real separable Hilbert space endowed with its Borel σ-algebra $B_{\mathscr{H}}$. Hilbert space is considered for it eases the asymptotic theorems derivation and agrees with the empirically recovered curves from discrete observations, see the survey by Mas and Pumo (2011) and the references therein. A general FAR(1) model is defined as

$$X_t - \mu = \rho(X_{t-1} - \mu) + \varepsilon_t, \tag{4.1}$$

where X_t takes values in $\mathscr{H}$ at time point t, and μ is the mean function. The operator ρ is a bounded linear operator from $\mathscr{H}$ to $\mathscr{H}$. The independent identically distributed sequence $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a strong $\mathscr{H}$-white noise with zero mean and $0 < E\|\varepsilon_0\|^2 = \cdots = E\|\varepsilon_t\|^2 < \infty$. The norm $\|\cdot\|$ is induced from the inner product $\langle\cdot,\cdot\rangle$ of $\mathscr{H}$. Bosq (2000) shows that a stationary solution exits for Equation (4.1) under the condition $\|\rho^j\|_{\mathscr{L}} < 1$ for $j \geq 1$. Here $\|\cdot\|_{\mathscr{L}}$ is the norm defined in $\mathscr{L}$, which is the space of bounded linear operators from $\mathscr{H}$ to $\mathscr{H}$.

In our study, the convolution kernel operator is particularly considered due to

its popularity in literature and the FAR(1) model becomes

$$X_t(\tau) - \mu(\tau) = \int_0^1 K(\tau - s)[X_{t-1}(s) - \mu(s)]ds + \varepsilon_t(\tau), \quad \tau \in [0,1], \tag{4.2}$$

where $X_t(\tau) \in C_{[0,1]}$ and is now a real continuous function defined on $[0,1]$. We have for the kernel function $K \in L^2([0,1]), \|K\|_2 < 1$ and K is 1-periodic, where $\|\cdot\|_2$ denotes the L^2 norm in $C_{[0,1]}$. The FAR model with convolution kernel as defined in Equation (4.2) can be considered as a natural generalization from discrete VAR modeling to FAR modeling which is continuous, and simultaneously a smooth transition from finite to infinite cases by adopting the integral format. Another reason that motivates us to define ρ as a convolution kernel operator is that it belongs to the class of Hilbert-Schmidt operators, which can be represented by functions. Compared to the general operator in (4.1), this advantage of Hilbert-Schmidt operator makes the modeling and estimation implementable in applications; while the implementation can be hardly expected for the abstract general operator. Details of definition of Hilbert-Schmidt operator and its norm are discussed later in Section 4.1.2.

As mentioned before, most existing works estimate the model parameter by YW approach under the assumption that the data are generated by a stationary FAR process. To deal with the existence of non-stationarity, we use ML estimation approach, in which case the likelihood function is an essential part in the adaptive procedure of AFAR modeling. Related details are elaborated in later sections.

However, ML estimation often fails in infinite-dimensional parameter spaces, see Grenander (1981). Based on Grenander's theory of sieves, Mourid and Bensmain (2006) derive a consistent ML estimator in FAR(1) modeling framework, where a likelihood function is defined over an approximating subspace of the original parameter space. We refer to Geman and Hwang (1982), Chen (2007) for more theoretical results on sieve estimation. As the primary interest is to estimate the FAR operator under stationarity in the aforementioned studies, the mean function is usually ignored and set to zero, which eases theoretical derivations but limits statistical inferences. In this section, we define a new likelihood function based on the method of sieves that allows the estimation not only for the FAR operator, but also for the mean function and the stochastic variations. We also develop the consistency results of the sieve estimators.

4.1.1 Fourier basis expansion and sieve estimation

We expand all the functions by using the trigonometric basis functions in $L^2([0,1])$:

$$\phi_0 = 1_{[0,1]}, \quad \phi_{2k}(\tau) = \sqrt{2}\cos 2\pi k\tau, \quad \phi_{2k-1}(\tau) = \sqrt{2}\sin 2\pi k\tau,$$

where ϕ_0, ϕ_{2k} and ϕ_{2k-1} are orthonormal. The functions in Equation (4.2) can be represented by Fourier basis functions and coefficients as follows:

$$\begin{aligned} X_t(\tau) &= a_{t,0} + \sum_{k=1}^{m_n}[b_{t,k}\phi_{2k-1}(\tau) + a_{t,k}\phi_{2k}(\tau)], \\ \varepsilon_t(\tau) &= a_0(\varepsilon_t) + \sum_{k=1}^{m_n}[b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau)], \\ K(\tau) &= c_0 + \sum_{k=1}^{m_n}[d_k\phi_{2k-1}(\tau) + c_k\phi_{2k}(\tau)], \end{aligned} \tag{4.3}$$

where $a_{t,0}, a_{t,k}, b_{t,k}$ are the Fourier coefficients of the constant, cosine, and sine basis functions respectively for the observed functional data X_t; and $a_0(\varepsilon_t), a_k(\varepsilon_t), b_k(\varepsilon_t)$ are the Fourier coefficients for the unknown innovations ε_t, while c_0, c_k, d_k are for the unknown kernel function K. Here we assume the Fourier coefficients $a_0(\varepsilon_t), a_k(\varepsilon_t)$ and $b_k(\varepsilon_t)$ are independent and identically Gaussian distributed with mean zero and variance σ_k^2. This assumption agrees with the multivariate normal distribution assumption for the density function, which is defined later.

In our study, Fourier basis functions are considered. By the computational properties of Fourier basis functions, it is much easier, compared to other basis functions such as splines and wavelets, to derive the density, likelihood functions and the estimators in closed-form expressions with the appearance of the basis coefficients only. Without the close-formed derivation, we need to handle additional difficulties in the proof of consistency by using splines or wavelets bases. In addition, the periodicity of the electricity price data also serves as a motivation

of adopting Fourier basis functions in real data analysis. However, it is worth to mention that wavelets have developed tremendously in lots of areas of applications. Compared to Fourier basis functions, wavelets are well localized in both frequency and time domains and especially perform well on tracking sharp or highly localized data features. Future works with applications by wavelets will be studied; but in this thesis, we will focus our analysis on the framework expanded by Fourier basis functions first.

At this moment, for the notation simplicity, we ignore the mean function and assign it to be zero first. In addition, the kernel function K is assumed to be an even function with all the sine coefficients d_k being zero. A sieve is a sequence of subsets of the infinite parameter space. We denote the sequence of subsets by $\{\Theta_{m_n}\}$ with $\Theta_{m_n} \subseteq \Theta_{m_n+1}$ and the union of the subsets $\bigcup \Theta_{m_n}$ is dense in the parameter space. The parameter estimation and the optimization of the likelihood function will be conducted on the subset of the parameter space, allowing that the dimension of the subset increases as the sample size increases. In our study, the approximating subspace or the sieve $\{\Theta_{m_n}\}$ is defined as

$$\Theta_{m_n} = \Big\{ K \in L^2 | K(\tau) = c_0 + \sum_{k=1}^{m_n} c_k \sqrt{2} \cos 2\pi k\tau, \tau \in [0,1], \sum_{k=1}^{m_n} k^2 c_k^2 \leq \nu m_n \Big\}, \qquad (4.4)$$

where $m_n \to +\infty$ as $n \to +\infty$ and ν is a given positive constant such that the constraint for c_k is generally satisfied without any sacrifice of the growth rate of m_n. It is worth to mention that there are alternative choices for the coefficients

constraint for sieves, see Geman and Hwang (1982) and Chen (2007). Here the constraint with quadratic forms of c_k is used mainly because the unknown function to be estimated is a continuous function with L^2 norm and the quadratic form of the constraint agrees with the quadratic likelihood function, which is derived later.

Therefore, Equation (4.2) can be rewritten as

$$\begin{aligned} X_t(\tau) &= \int_0^1 \left\{ c_0 + \sum_{k=1}^{m_n} c_k \phi_{2k}(\tau - s) \right\} \\ &\quad \times \left\{ a_{t-1,0} + \sum_{k=1}^{m_n} b_{t-1,k}\phi_{2k-1}(s) + \sum_{k=1}^{m_n} a_{t-1,k}\phi_{2k}(s) \right\} ds \\ &\quad + a_0(\varepsilon_t) + \sum_{k=1}^{m_n} [b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau)]. \end{aligned} \tag{4.5}$$

By the properties of Fourier basis functions, the above Equation (4.5) can be arranged as follows:

$$\begin{aligned} X_t(\tau) &= c_0 a_{t-1,0} + \sum_{k=1}^{m_n} \frac{1}{\sqrt{2}} c_k b_{t-1,k}\phi_{2k-1}(\tau) + \sum_{k=1}^{m_n} \frac{1}{\sqrt{2}} c_k a_{t-1,k}\phi_{2k}(\tau) \\ &\quad + a_0(\varepsilon_t) + \sum_{k=1}^{m_n} [b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau)] \\ &= a_{t,0} + \sum_{k=1}^{m_n} [b_{t,k}\phi_{2k-1}(\tau) + a_{t,k}\phi_{2k}(\tau)]. \end{aligned}$$

It provides the following relationship of the Fourier coefficients:

$$\begin{aligned} a_{t,0} &= c_0 a_{t-1,0} + a_0(\varepsilon_t), \\ a_{t,k} &= \frac{1}{\sqrt{2}} c_k a_{t-1,k} + a_k(\varepsilon_t), \\ b_{t,k} &= \frac{1}{\sqrt{2}} c_k b_{t-1,k} + b_k(\varepsilon_t). \end{aligned}$$

For a general case with the mean function being non-zero, we have

$$X_t(\tau) = \mu(\tau) - \int_0^1 K(\tau - s)\mu(s)ds + \int_0^1 K(\tau - s)X_{t-1}(s)ds + \varepsilon_t(\tau).$$

Suppose p_0, p_k, q_k are assigned as the Fourier coefficients corresponding to the constant, cosine and sine basis functions respectively for the function $\mu(\tau) - \int_0^1 K(\tau - s)\mu(s)ds$. We obtain a general relationship for the Fourier coefficients:

$$\begin{aligned} a_{t,0} &= p_0 + c_0 a_{t-1,0} + a_0(\varepsilon_t), \\ a_{t,k} &= p_k + \frac{1}{\sqrt{2}} c_k a_{t-1,k} + a_k(\varepsilon_t), \\ b_{t,k} &= q_k + \frac{1}{\sqrt{2}} c_k b_{t-1,k} + b_k(\varepsilon_t). \end{aligned}$$

For notational simplicity, we denote the unknown parameters by $\theta = (c_0, c_k, p_0, p_k, q_k, \sigma_0, \sigma_k)$ for $k = 1, \cdots, m_n$ with $m_n \to +\infty$ as $n \to +\infty$ in later sections.

On each finite parameter subspace, we define a transition density under the multivariate normal distributional assumption such that

$$\begin{aligned} g(X_t, X_{t-1}, \rho) &= \frac{2\pi^{-(2m_n+1)/2}}{\sigma_0 \prod_{k=1}^{m_n} \sigma_k^2} \cdot \exp\Big\{ - \frac{1}{2\sigma_0^2}(a_{t,0} - p_0 - c_0 a_{t-1,0})^2 \\ &\quad - \sum_{k=1}^{m_n} \frac{1}{2\sigma_k^2}\Big[(b_{t,k} - q_k - \frac{1}{\sqrt{2}} c_k b_{t-1,k})^2 + (a_{t,k} - p_k - \frac{1}{\sqrt{2}} c_k a_{t-1,k})^2\Big]\Big\}, \end{aligned}$$

and the conditional log-likelihood $L(X_1, ..., X_n; \rho)$ is

$$\begin{aligned} L(X_1, ..., X_n; \rho) &= \log\Big\{ \prod_{t=2}^{n} g(X_t, X_{t-1}, \rho)\Big\} \\ &= -\frac{(2m_n + 1)(n-1)}{2} \log 2\pi - (n-1)\log\sigma_0 - (n-1)\sum_{k=1}^{m_n} \log\sigma_k^2 \end{aligned}$$

$$
\begin{aligned}
&- \quad \frac{1}{2\sigma_0^2}\sum_{t=2}^{n}(a_{t,0}-p_0-c_0a_{t-1,0})^2 \\
&- \quad \sum_{t=2}^{n}\sum_{k=1}^{m_n}\frac{1}{2\sigma_k^2}\Big\{(b_{t,k}-q_k-\frac{1}{\sqrt{2}}c_kb_{t-1,k})^2 \\
&+ \quad (a_{t,k}-p_k-\frac{1}{\sqrt{2}}c_ka_{t-1,k})^2\Big\}.
\end{aligned}
$$

By maximizing the log-likelihood on the sieve $\{\Theta_{m_n}\}$, we obtain the sieve estimators for kernel function K:

$$
\begin{aligned}
\tilde{c}_0 &= \frac{\sum_{t=2}^{n}a_{t,0}\sum_{t=2}^{n}a_{t-1,0}-(n-1)\sum_{t=2}^{n}a_{t,0}a_{t-1,0}}{(\sum_{t=2}^{n}a_{t-1,0})^2-(n-1)\sum_{t=2}^{n}a_{t-1,0}^2}, \qquad (4.6)\\
\tilde{c}_k &= \sqrt{2}\frac{\sum_t(a_{t,k}a_{t-1,k}+b_{t,k}b_{t-1,k})-(\sum_t a_{t,k}\sum_t a_{t-1,k}+\sum_t b_{t,k}\sum_t b_{t-1,k})/(n-1)}{\sum_t(a_{t-1,k}^2+b_{t-1,k}^2)-\{(\sum_t a_{t-1,k})^2+(\sum_t b_{t-1,k})^2\}/(n-1)}.
\end{aligned}
$$

Now the kernel K can be obtained via estimating its Fourier coefficients. Our developed ML estimators are different from those in Mourid and Bensmain (2006) as the transition density function differs. By using the new transition density, we can derive the estimators of the innovation variation σ_k^2 as well as the intercept term:

$$
\tilde{\sigma}_0^2=\frac{\sum_{t=2}^{n}(a_{t,0}-\tilde{p}_0-\tilde{c}_0a_{t-1,0})^2}{n-1}, \quad \tilde{p}_0=\frac{-\tilde{c}_0\sum_{t=2}^{n}a_{t-1,0}+\sum_{t=2}^{n}a_{t,0}}{n-1},
$$

$$
\tilde{\sigma}_k^2=\frac{\sum_{t=2}^{n}[(b_{t,k}-\tilde{q}_k-\frac{1}{\sqrt{2}}\tilde{c}_kb_{t-1,k})^2+(a_{t,k}-\tilde{p}_k-\frac{1}{\sqrt{2}}\tilde{c}_ka_{t-1,k})^2]}{2(n-1)},
$$

$$
\tilde{p}_k=\frac{\sum_{t=2}^{n}a_{t,k}-\frac{1}{\sqrt{2}}\tilde{c}_k\sum_{t=2}^{n}a_{t-1,k}}{n-1}, \quad \tilde{q}_k=\frac{\sum_{t=2}^{n}b_{t,k}-\frac{1}{\sqrt{2}}\tilde{c}_k\sum_{t=2}^{n}b_{t-1,k}}{n-1}.
$$

Given a fixed m_n, the estimator $\tilde{\sigma}_k$ and all the $\tilde{c}_k$, we can calculate the fitted log-likelihood $L(X_1,...,X_n;\tilde{\rho})$.

4.1.2 Consistency results for sieve estimators

Now we derive the consistency results of the sieve estimators. Let $H(\rho, \beta)$ denote the conditional entropy between operator ρ and another given operator β:

$$H(\rho, \beta) = E_\rho \left[\log g(X_t, X_{t-1}, \beta)\right],$$

In the following theorems and the proof in Appendix, $H(\rho, \rho) - H(\rho, \beta)$ or its analogue forms appear and are calculated. It is worth to mention that in our framework, $H(\rho, \rho) - H(\rho, \beta)$ is just the Kullback-Leibler information for two multivariate normal distributions with different sets of parameters corresponding to operators ρ and β.

To determine the rate of growth size m_n, we consider the following two conditions.

C1: If there exits a sequence $\{\rho_{m_n}\}$ such that $\rho_{m_n} \in \Theta_{m_n} \forall n$ and $H(\rho_{0|\Theta_{m_n}}, \rho_{m_n}) \to H(\rho_{0|\Theta_{m_n}}, \rho_{0|\Theta_{m_n}})$, then $\|\rho_{m_n} - \rho_{0|\Theta_{m_n}}\|_{\mathscr{S}} \to 0$. Here $\rho_{0|\Theta_{m_n}}$ denotes the projection of the true operator ρ_0 on the sieve Θ_{m_n}.

C2: There exits a sequence $\rho_{m_n} \in \Theta_{m_n}$ as described in C1 such that $H(\rho_{0|\Theta_{m_n}}, \rho_{m_n}) \to H(\rho_{0|\Theta_{m_n}}, \rho_{0|\Theta_{m_n}})$.

The norm $\|\cdot\|_{\mathscr{S}}$ is a Hilbert-Schmidt norm for the convolution kernel operator. Recall that a linear operator ρ on a Hilbert space $\mathscr{H}$ with norm $\|\cdot\|$ and inner

product $\langle \cdot, \cdot \rangle$ is Hilbert-Schmidt if $\rho(\cdot) = \sum_j \lambda_j \langle \cdot, e_j \rangle f_j$, where $\{e_j\}$ and $\{f_j\}$ are orthonormal bases of $\mathscr{H}$ and $\{\lambda_j\}$ is a real sequence such that $\sum_j \lambda_j^2 < \infty$. The convolution kernel operator defined in Section 4.1 satisfies the definition and its Hilbert-Schmidt norm is $\|\rho\|_{\mathscr{S}} = (\sum_j \lambda_j^2)^{1/2}$. The Hilbert-Schmidt norm is chosen for our study because of the fact that the convolution kernel operator defined in this thesis forms a class of operators embedded in the whole space of Hilbert-Schmidt operators and for any convolution kernel operator ρ, we have the Hilbert-Schmidt norm of ρ equal to the L^2 norm of its kernel function, that is $\|\rho\|_{\mathscr{S}} = \|K\|_2$.

Theorem 4.1.1. *Assume $\{\Theta_{m_n}\}$ is chosen such that condition C1 and C2 are in force. Suppose that for each $\delta > 0$, we can find subsets $\Gamma_1, \Gamma_2, \cdots, \Gamma_{l_{m_n}}$ of $\Theta_{m_n}, m_n = 1, 2, ...$ such that*

(i) $D_{m_n} \subseteq \bigcup_{k=1}^{l_{m_n}} \Gamma_k$, where $D_{m_n} = \{\rho \in \Theta_{m_n} | H(\rho_{0|\Theta_{m_n}}, \rho) \leq H(\rho_{0|\Theta_{m_n}}, \rho_{m_n}) - \delta\}$ for every $\delta > 0$ and every m_n.

(ii) $\sum_{n=1}^{+\infty} l_{m_n} (\varphi_{m_n})^n < +\infty$, where given l sets $\Gamma_1, ..., \Gamma_l$ in Θ_{m_n}, φ_{m_n} is defined as $\varphi_{m_n} = \sup_k \inf_{t \geq 0} E_{\rho_{0|\Theta_{m_n}}} \exp\left[t \log\left\{\frac{g(X_i, X_{i-1}, \Gamma_k)}{g(X_i, X_{i-1}, \rho_{m_n})}\right\}\right]$.

Then we have $\sup_{\hat{\rho}_n \in M_m^n} \|\hat{\rho}_n - \rho_{0|\Theta_{m_n}}\|_{\mathscr{S}} \to 0$ a.s.

Note that in Theorem 4.1.1, $g(X_i, X_{i-1}, \Gamma_k) = \sup_{\beta \in \Gamma_k} g(X_i, X_{i-1}, \beta)$. The set M_m^n is defined as a set of all the ML estimators on Θ_{m_n} given sample size n, i.e., $M_m^n = \left\{\rho \in \Theta_{m_n} | L(X_1, ..., X_n; \rho) = \sup_{\beta \in \Theta_{m_n}} L(X_1, ...X_n; \beta)\right\}$. We follow Mourid

and Bensmain (2006) for the proof of Theorem 4.1.1 to show that the ML estimator converges to $\rho_{0|\Theta_{m_n}}$, the projection of the true operator on sieve. Together with the convergence of $\rho_{0|\Theta_{m_n}}$ to the true operator ρ_0 as the sieve dimension grows, we obtain the result that the ML estimator converges to the true operator ρ_0. An application of Theorem 4.1.1 provides the growth rate of m_n and the almost sure convergence of the sieve estimator when the kernel function K satisfying (4.4) is considered in FAR modeling.

Theorem 4.1.2. *If* $m_n = o\ (n^{1/3-\eta})$ *for* $\eta > 0$, *then* $\|\hat{K}_{m_n} - K_{0|\Theta_{m_n}}\|_2 \to 0$ *a.s. when* $n \to +\infty$ *and* $\|\cdot\|_2$ *is* L^2 *norm in* $C_{[0,1]}$.
$\hat{K}_{m_n}$ *is the sieve estimator on* Θ_{m_n} *and* $K_{0|\Theta_{m_n}}$ *is the projection of the true kernel function* K_0 *on* Θ_{m_n}.

The proof of Theorem 4.1.2 can be achieved by checking the conditions of Theorem 4.1.1. Similar arguments can be followed as shown in Mourid and Bensmain (2006). However, as the density function $g(X_t, X_{t-1}, \rho)$ with the corresponding kernel function K has been modified in our study, we will show the different parts from the arguments in the aforementioned paper. Proof is detailed in Appendix. As $n, m_n \to \infty$, we have $\|K_{0|\Theta_{m_n}} - K_0\|_2 \to 0$ because $K_{0|\Theta_{m_n}}$ is just the Fourier truncation of the true kernel K_0 on Θ_{m_n} and finally we have the sieve estimator $\hat{K}_{m_n}$ converges to the true kernel function K_0.

The ML estimation is consistent under stationarity. Non-stationarity or near non-stationarity is a stylized fact of many time series data such as electricity price curves, yield curves and weather curves, etc. Even though there are few frequent changes in these data, it poses many challenges, not only for theoretical modeling, making inferences and conducting tests, but also for real time forecasting. We will discuss how to estimate and forecast non-stationary functional data in next section.

4.2. AFAR modeling under non-stationarity

In this section, we propose an AFAR modeling to estimate the joint dynamics of functional data. The AFAR model of order 1 is defined as follows:

$$X_t - \mu_t = \rho_t(X_{t-1} - \mu_{t-1}) + \varepsilon_t, \tag{4.7}$$

where the mean function, FAR operator and stochastic variations are allowed to be time dependent. Under stationarity that all the parameters are constant, the AFAR model coincides with the stationary FAR model. For non-stationary processes, the AFAR modeling is flexible without explicitly assuming any change type and timing. In either case, we adaptively estimate the time-varying model under the local homogeneity assumption. Local homogeneity assumes that at any particular forecast origin, say e.g., n, there exists a past time interval $I_n = [n - \ell_n + 1, n]$,

over which the local sample can be well approximated by an FAR(1) model with constant parameters, and $\mu_s \approx \mu_n$, $\rho_s \approx \rho_n$, $\varepsilon_s \approx \varepsilon_n$ for all $s \in I_n$. A sequential testing procedure is used to find the longest interval that satisfies the local homogeneity assumption. Over the local intervals, the AFAR model with time-varying parameters does not deviate much from the FAR model with constant ones. Thus the ML estimation can be safely conducted. Like in a rolling window technique, the AFAR model is recursively estimated in a local window, but now with a changing window size ℓ_n.

In our study, we only consider the AFAR model of order 1. Although it would be interesting to explore non-stationary FAR models of higher orders, the selection of lag order 1 is motivated by the fact that FAR models can be reduced to FAR(1) by using a suitable Markov representation, see Bosq (2000). In addition, the adoption of the simplest model structure is significantly motivated by the tractability of easy interpretation and good out-of-sample forecast performance. Thus we opt for the AFAR(1) model, but pay attention on a careful selection of the local interval of homogeneity.

Now we relax the assumption of stationarity to the local homogeneity assumption. Given a sample of random curves $X_1, ..., X_n$, our interest is to identify a local interval, over which the local sample is represented by a stationary FAR(1) model.

The estimation bias occurs as we use the misspecified stationary model for non-stationary data. Therefore, the interval selection should lead to small estimation bias in modeling, and simultaneously reach to the highest information efficiency by using as many data as possible. Such an interval is called interval of local homogeneity. By fitting the FAR model over the interval of local homogeneity, we can reach to the best estimation accuracy with a small modeling bias. The small modeling bias condition will be introduced and discussed in Section 4.2.3. The fitted model is further used to forecast the $h-$step ahead curve $X_{n+h}(\tau)$ at the forecast origin n. At an arbitrary forecast origin n, suppose the interval of local homogeneity $I_n = [n - \ell_n + 1, n]$ is given. We consider ℓ_n past curves $X_{n-\ell_n+1}, ..., X_n$, whose coefficients $\{a_{t,0}\}_{t=n-\ell_n+1}^{n}$ and $\{a_{t,k}, b_{t,k}\}_{t=n-\ell_n+1}^{n}$ are obtained by the Fourier Expansion (4.3). The local log-likelihood defined on the sieve is

$$\begin{aligned} L(I_n;\theta_n) &= -\frac{(2m_n+1)(\ell_n-1)}{2}\log 2\pi - (\ell_n-1)\log\sigma_{n,0} - (\ell_n-1)\sum_{k=1}^{m_n}\log\sigma_{n,k}^2 \\ &- \frac{1}{2\sigma_{n,0}^2}\sum_{t=n-\ell_n+2}^{n}(a_{t,0} - p_{n,0} - c_{n,0}a_{t-1,0})^2 \\ &- \sum_{t=n-\ell_n+2}^{n}\sum_{k=1}^{m_n}\frac{1}{2\sigma_{n,k}^2}\Big[(b_{t,k} - q_{n,k} - \frac{1}{\sqrt{2}}c_{n,k}b_{t-1,k})^2 \\ &+ (a_{t,k} - p_{n,k} - \frac{1}{\sqrt{2}}c_{n,k}a_{t-1,k})^2\Big], \end{aligned}$$

where $\theta_n = (c_{n,0}, c_{n,k}, p_{n,0}, p_{n,k}, q_{n,k}, \sigma_{n,0}, \sigma_{n,k})$ for $k = 1, \cdots, m_n$. The respective local ML estimator, denoted as $\tilde{\theta}_n$, is $\tilde{\theta}_n = \arg\max L(I_n;\theta_n)$. In the following, we will use $L(I_n;\theta_n)$ to denote the log-likelihood function on local interval I_n, with ρ_n

being the corresponding operator to θ_n.

4.2.1 Adaptive estimation procedure

In practice, the interval of local homogeneity is unknown. The question is how to identify it or equivalently how to select interval length ℓ_n at time point n. With too large a value, there is a high probability of having a non-trivial modeling bias, which violates the local homogeneity assumption. On the contrary, a small value of ℓ_n though satisfying a small modeling bias, unnecessarily discards too many observations that are useful for estimation. The goal is to select the longest interval that does not violate the local homogeneity assumption.

Suppose there are S candidate intervals at time point n, which contain the interval of local homogeneity I_n, such that

$$I_n^{(1)}, \cdots, I_n^{(S)} \quad \text{with} \quad I_n^{(1)} \subset \cdots \subset I_n^{(S)}.$$

A sequential testing procedure helps to select the longest interval that satisfies the local homogeneity. Beyond the selected interval, there would be a high probability of structural changes. The adaptive technique proceeds as follows. The procedure starts from the shortest interval $I_n^{(1)}$, over which the local homogeneity assumption is accepted by default. We accept the interval and denote the accepted estimator by $\hat{\theta}_n = \widetilde{\theta}_n^{(1)}$. Iteratively, the procedure extends to the next interval $I_n^{(s)}$ for $s = 2$,

where we introduce the test statistic

$$T_n^{(s)} = \left| L(I_n^{(s)}; \tilde{\theta}_n^{(s)}) - L(I_n^{(s)}; \hat{\theta}_n) \right|^r, \tag{4.8}$$

where r is a given positive constant and we take $r = 1/2$ in our study. The test statistic measures the divergence of the hypothetical constant FAR model from the recently accepted (data-driven) local model. If the divergence is significant, where the significance level is controlled by a critical value ζ_s, it indicates that there is a significant structure change larger than the one arising due to sampling changes. In this case, we reject the null hypothesis of local homogeneity and terminate the selection procedure. The last accepted interval would be the final selection . Otherwise, we accept the longer interval $I_n^{(s)}$, and update the estimate $\hat{\theta}_n = \widetilde{\theta}_n^{(s)}$. The test procedure is continued for $s > 2$ until either a change is detected or the longest candidate interval is reached.

The sequential testing procedure can be carried out for any particular forecast origin n. The algorithm is formulated as follows:

(1) $I_n^{(1)}$ is accepted by default. We initialize the adaptive estimator $\hat{\theta}_n = \tilde{\theta}_n^{(1)}$.

(2) In the subsequent steps s, we consider $I_n^{(s)}$, with $2 \leq s \leq S$:

- if $T_n^{(s)} \leq \zeta_s$, we accept $I_n^{(s)}$. Update the adaptive estimator $\hat{\theta}_n = \tilde{\theta}_n^{(s)}$ and set $s = s + 1$.
- otherwise we reject the null hypothesis and terminate the procedure.

The last considered interval candidate $I_n^{(s-1)}$ is the final selection. We set $\hat{\theta}_n = \tilde{\theta}_n^{(s-1)}$.

(3) If the procedure is never terminated before the longest candidate interval $I_n^{(S)}$ is reached, the final adaptive estimator is $\hat{\theta}_n = \tilde{\theta}_n^{(S)}$.

The critical values $\zeta_1, \cdots, \zeta_S$ are crucial in the adaptive estimation. As the sampling distribution of the test statistic is unknown, we show how to calibrate the critical values via Monte Carlo simulations in next section.

4.2.2 Critical value calibration

The success of the adaptive selection procedure depends on the critical values, $\zeta_1, \cdots, \zeta_S$, which are calibrated in Monte Carlo experiments. As the critical values control the significance level under the local homogeneity assumption that requests a small modeling bias, we generate samples with homogeneity and measure the modeling bias using an adaptive estimation. The critical values are selected such that they are capable of providing the prescribed performance of the testing procedure.

In particular, we generate homogeneous FAR data processes according to the Fourier coefficients' relationship:

$$a_0(X_t) \quad = \quad p_0^* + c_0^* a_0(X_{t-1}) + a_0(\varepsilon_t)$$

$$\begin{aligned} a_k(X_t) &= p_k^* + \frac{1}{\sqrt{2}} c_k^* a_k(X_{t-1}) + a_k(\varepsilon_t) \\ b_k(X_t) &= q_k^* + \frac{1}{\sqrt{2}} c_k^* b_k(X_{t-1}) + b_k(\varepsilon_t). \end{aligned}$$

We denote the constant parameters as $\theta^* = (c_0^*, c_k^*, p_0^*, p_k^*, q_k^*, \sigma_0^*, \sigma_k^*)$ for $k = 1, \cdots, m_n$ and $t = 1, \cdots, T$. The size T of the generated stationary data is the same as the length of the longest candidate interval $I_n^{(S)}$ as proposed in Section 4.2.1. In this section, for simplicity, we denote the candidate intervals by $I^{(1)}, \cdots, I^{(S)}$ without any subscripts since we will stay at the last time point T and look back with past data observations for the calibration of critical values. The generated stationarity can be considered as a special case of local homogeneity. Thus we can use both methods, the ML estimation and the adaptive estimation given the critical values. The critical values are selected such that the divergence between the ML estimation and adaptive estimation is bounded above by a small value, which will be defined in (4.10).

The ML estimation can be safely conducted over every interval $I^{(s)}$, which gives the ML estimator $\tilde{\theta}^{(s)}$. Especially, the estimation error can be measured by

$$\mathfrak{R}_s = E_{\theta^*} \left| L(I^{(s)}; \tilde{\theta}^{(s)}) - L(I^{(s)}; \theta^*) \right|^r, \tag{4.9}$$

where $\mathfrak{R}_s$ can be computed numerically with the knowledge of θ^* and r, for $s = 1, \cdots, S$. For the adaptive estimation, once a set of critical values $\zeta_1, \cdots, \zeta_S$ is given, we can employ the adaptive procedure by checking the significance of the

test statistic as described in Section 4.2.1, to obtain the adaptive estimate $\hat{\theta}^{(s)}$ corresponding to interval $I^{(s)}$ and critical values ζ_s. The difference from Section 4.2.1 is that now the adaptive estimation is conducted on the generated stationary data with underlying parameter θ^* and the forecast origin is the last time point T. The temporal divergence between the constant FAR model and the AFAR model, denoted as $D^{(s)}$, can be measured as follows:

$$D^{(s)} = \left|L(I^{(s)}; \tilde{\theta}^{(s)}) - L(I^{(s)}; \hat{\theta}^{(s)})\right|^r.$$

The adaptive estimation should behave as well as the true underlying characteristics under the null of time homogeneity, in the sense that the modeling bias is bounded by the estimation error with the knowledge of the true model, i.e.,

$$E_{\theta^*}\left[D^{(s)}\right] = E_{\theta^*}\left|L(I^{(s)}; \tilde{\theta}^{(s)}) - L(I^{(s)}; \hat{\theta}^{(s)})\right|^r \leq \alpha\mathfrak{R}, \tag{4.10}$$

where $\mathfrak{R} = \max_{s \leq S} \mathfrak{R}_s$. The parameter α corresponds to the sensitivity of the adaptive model and the choice of α value is usually subjective and selected by experience. A small value of α will result in large critical values which eventually lead to small divergence and make it easy to accept the null hypothesis over long interval candidates, but insensitive to detect possible changing parameters; while a large α value comes with small critical values leading to a relatively more stringent test that unnecessarily favors short intervals and thus discards useful past observations and results in higher parameter uncertainty. We suggest choosing the possibly smallest critical values that either satisfy the risk bound everywhere or result in

the best forecast accuracy. It appears that such a choice is not only attainable in this artificial stationarity, but also leads to comparable forecast performance in a general case with parameter changes, as will be shown in Section 4.3 of simulation study and in Section 4.4 of real data analysis.

The only unknown parameters in (4.10) are the critical values, which can be calibrated. Here we will elaborate the procedure according to which the critical values are calibrated. The parameter r is fixed as $1/2$ to provide a stable performance of estimation and simulation, see Belomestny and Spokoiny (2007). As mentioned before, the sequential testing procedure in Section 4.2.1 is adopted to select the critical values. At step $s = 1$, we set the first critical value $\zeta_1 = \infty$ such that the homogeneity of the shortest interval $I^{(1)}$ is guaranteed and $\hat{\theta} = \widetilde{\theta}^{(1)}$ is assigned by default. For selection of ζ_2, we set $\zeta_s = \infty$ for all $s \geq 3$ to emphasize the contribution of ζ_2. The value of ζ_2 is selected as the minimum value such that the following risk inequality is satisfied:

$$E_{\theta^*}\big|L(I^{(s)};\tilde{\theta}^{(s)}) - L(I^{(s)};\hat{\theta}^{(s)})\big|^{1/2} \leq \frac{1}{S-1}\mathfrak{R}, \quad s = 2, \cdots, S.$$

Subsequently, with $\zeta_1, \zeta_2, \cdots, \zeta_{s-1}$ calculated and fixed, the value of ζ_s for $s = 3, \cdots, S$ is selected as the minimum value that satisfies:

$$E_{\theta^*}\big|L(I^{(j)};\tilde{\theta}^{(j)}) - L(I^{(j)};\hat{\theta}^{(j)})\big|^{1/2} \leq \frac{s-1}{S-1}\mathfrak{R}, \quad j = s, \cdots, S.$$

It is worth to mention that the weight $(s-1)/(S-1)$ for $s = 2, \cdots, S$ is introduced

to replace α in the risk bound, which reflects and deals with the increased bias as the sample size increases.

The computation of critical values relies on a set of hyperparameters (S and θ^*). The parameter θ^* is used in the Monte Carlo generation. It should be close to the true parameter underlying the real data series. In the numerical analysis, we consider using the ML estimator before the first forecast origin and adopt the calibrated critical values everywhere. The adaptive estimation is found to be robust to the selection of the hyperparameters, as will be shown in Section 4.3.

4.2.3 Theoretical properties for the adaptive estimator

In this section, we study the theoretical results describing the accuracy of the adaptive estimation when the underlying data process is locally approximated by a stationary model with a constant parameter. Besides, the small modeling bias (SMB) condition is introduced and discussed. We refer to Spokoiny (2009) for the theoretical results of the univariate adaptive models.

As discussed, the AFAR modeling parameter is time-varying and we denote by θ_t for all points t in an interval I. In our study, we assume that the time-varying parameter θ_t can be well approximated by a constant θ for all $t \in I$. To measure the divergence of the approximation of θ_t by using a constant parameter θ on a

local interval I, we define

$$\Delta_I(\theta) = \sum_{t \in I} \mathscr{K}\big(\theta_t, \theta\big),$$

where $\mathscr{K}(\theta, \theta')$ is the Kullback-Leibler information for two multivariate normal distributions with parameters θ and θ'.

Remark 4.2.1. *Note that the Kullback-Leibler information is defined to be* $\mathscr{K}(\theta, \theta') = E_\theta\left[\log(\frac{dP_\theta}{dP_{\theta'}})\right]$, *where* P_θ *and* $P_{\theta'}$ *are the probability measures with respect to parameters* θ *and* θ' *respectively.*

Now we introduce the SMB condition, which assumes that for some θ in the parameter space, $\Delta_I(\theta)$ is bounded by a small constant with a high probability. In other words, it means the true time-varying model can be well approximated on the interval I by a parametric model with a constant parameter θ. Under SMB condition, the constant parameter θ is considered as the target of estimation, with $\tilde{\theta}$ being the sieve ML estimate obtained on interval I. We will study the estimation accuracy of the ML estimate with the SMB assumption satisfied.

More generally, let $\varrho(\hat{\theta}, \theta)$ be a loss function for an estimate $\hat{\theta}$ constructed from the observations on interval I. The corresponding risk under the parametric measure P_θ is defined as $\mathfrak{R}(\hat{\theta}, \theta) = E_\theta[\varrho(\hat{\theta}, \theta)]$.

Theorem 4.2.2. *For some* θ *in the parameter space and a constant* $\Delta \geq 0$, *let*

$$E[\Delta_I(\theta)] \leq \Delta. \tag{4.11}$$

Then it holds for any estimate $\hat{\theta}$ measurable with respect to $\mathscr{F}_I$ such that

$$E\left[\log(1+\varrho(\hat{\theta},\theta)/\mathfrak{R}(\hat{\theta},\theta))\right] \leq \Delta + 1,$$

where $\mathscr{F}_I$ is the filtration or σ-field generated by the observations on interval I.

Condition (4.11) is called the SMB assumption. If we formulate the loss function $\varrho(\tilde{\theta},\theta) = |L(I,\tilde{\theta}) - L(I,\theta)|^r$ with the ML estimate $\tilde{\theta}$, we will obtain the following result.

Corollary 4.2.3. *Suppose the SMB condition (4.11) holds for some interval I and θ in the parameter space. Then for any $r > 0$, we have*

$$E\left[\log\left(1+|L(I;\tilde{\theta}) - L(I;\theta)|^r/\mathfrak{R}(\theta)\right)\right] \leq \Delta + 1,$$

where $\mathfrak{R}(\theta)$ is the parametric risk for the associated loss function.

This result shows that the estimation loss $|L_I(\tilde{\theta},\theta)|^r = |L(I;\tilde{\theta}) - L(I;\theta)|^r$ normalized by its parametric risk is bounded by a constant, provided that $\Delta_I(\theta)$ is sufficiently small. The proof of Theorem 4.2.2 can be found in Appendix.

Recall that we have S candidate intervals $I^{(1)}, \cdots, I^{(S)}$ which contain the interval of local homogeneity. At step s of the adaptive estimation procedure, the interval $I^{(s)}$ is involved for the homogeneity test. In this case, the deviation of the underlying model from the constant model is measured by

$$\Delta_s(\theta) = \Delta_{I^{(s)}}(\theta) = \sum_{t \in I^{(s)}} \mathscr{K}(\theta_t, \theta).$$

We can see that the modeling bias $\Delta_s(\theta)$ increases with s. When $\Delta_s(\theta)$ increases to be very large, the risk bound obtained in Corollary 4.2.3 explodes and becomes meaningless. We suppose that there exits an optimal s^* such that

$$E[\Delta_s(\theta)] \leq \Delta, \quad \text{for all } s \leq s^* \leq S, \tag{4.12}$$

for some $\Delta > 0$ and (4.12) does not hold for $s > s^*$. Now Theorem 4.2.2 implies the following result for the first s^* steps of the adaptive procedure.

Theorem 4.2.4. *Let θ and s^* be such that (4.12) holds for some $\Delta \geq 0$. Then*

$$E \log \Big(1 + \frac{|L_{I^{(s^*)}}(\tilde{\theta}^{(s^*)}, \theta)|^r}{\mathfrak{R}(\theta)}\Big) \leq 1 + \Delta,$$

$$E \log \Big(1 + \frac{|L_{I^{(s^*)}}(\tilde{\theta}^{(s^*)}, \hat{\theta}^{(s^*)})|^r}{\alpha\mathfrak{R}(\theta)}\Big) \leq 1 + \Delta,$$

where $\tilde{\theta}^{(s^*)}$ is the ML estimate for interval $I^{(s^*)}$ and $\hat{\theta}^{(s^*)}$ is the adaptive estimate obtained at step s^*, which also corresponds to interval $I^{(s^*)}$.

Theorem 4.2.4 shows that the distance between the ML estimate $\tilde{\theta}^{(s^*)}$ and θ is of the same magnitude as the distance between $\tilde{\theta}^{(s^*)}$ and $\hat{\theta}^{(s^*)}$. Therefore, we can conclude that the adaptive estimate $\hat{\theta}^{(s^*)}$ is close to the ML estimate $\tilde{\theta}^{(s^*)}$ and the adaptive estimation does not generate more errors into the ML estimation. This indicates the adaptive estimation for θ is accurate under the assumption of SMB. It is worth to mention that the risk bound results for Theorem 4.2.4 are derived for the general case when $r > 0$. For specific cases such as when $r = 1/2$, it is

possible the risk bound deviates from the results in Theorem 4.2.4. The derivation and discussion of the risk bound for the case when $r = 1/2$ have been studied. We refer to Spokoiny (2009) and Chen and Spokoiny (2013) for the result.

4.3. Simulation study

We conduct a simulation study to illustrate the finite sample properties of the AFAR modeling under both stationary and non-stationary situations. Furthermore, the robustness of the estimation performance for the adaptive procedure is investigated with respect to different choices of hyperparameters (θ^* and S). The results reveal that the AFAR modeling with time-dependent local window sizes introduces flexibility into the estimation, leading to a comparable performance with the true data generating process under a stationary scenario and generally reasonable accuracy under scenarios with regime shifts. Moreover, the analysis exhibits the robustness of the adaptive approach to the selection of different hyperparameters.

4.3.1 Stationarity: finite sample estimation accuracy

We first consider a stationary scenario with constant parameters, which is denoted as HOM. Functional data $\{X_t\}$ are generated according to the Fourier coefficients' relationship:

$$
\begin{aligned}
a_{t,0} &= p_0 + c_0 a_{t-1,0} + a_0(\varepsilon_t), \\
a_{t,k} &= p_k + \frac{1}{\sqrt{2}} c_k a_{t-1,k} + a_k(\varepsilon_t), \\
b_{t,k} &= q_k + \frac{1}{\sqrt{2}} c_k b_{t-1,k} + b_k(\varepsilon_t),
\end{aligned}
$$

where the innovations $a_k(\varepsilon_t)$ and $b_k(\varepsilon_t)$ are independent and identically Gaussian distributed with mean zero and constant variance σ_k^2 that are randomly generated from a uniform distribution $U[0,1]$. The other parameters p_0, p_k and q_k are randomly selected from $U[-1,1]$. To assure stationarity of the generated data, c_k's are randomly selected such that $c_0 \in [-1,1]$ and $c_k \in [-\sqrt{2},\sqrt{2}]$. Given the initial values $a_{0,0} = \frac{p_0}{1-c_0}$, $a_{0,k} = \frac{p_k}{1-c_k/\sqrt{2}}$ and $b_{0,k} = \frac{q_k}{1-c_k/\sqrt{2}}$, i.e., the mean of each AR process, we set $k = 1, \cdots, 4$ and generate 600 functional data processes with a constant parameter $\theta_0 = (c_0, c_k, p_0, p_k, q_k, \sigma_0, \sigma_k)$. The generation is repeated for 1000 times.

For each process, we use the past 300 observations to iteratively estimate the parameters, starting from $t = 301$ and ending at $t = 600$. The average values of the

ML estimators and their estimation error measurements, root mean squared error (RMSE) and mean of absolute deviation (MAD) are reported in Table 4.1, entitled as "true DGP (FAR)". Compared to the true parameters, the ML estimators are accurate and provide reasonable accuracy, with RMSE values smaller than 0.06 and MAD values smaller than 0.05.

To study the flexibility of the AFAR modeling under stationarity, we also implement the adaptive estimation to the generated homogeneous data. With the time-varying assumption, AFAR model in (4.7) is a misspecified model under stationarity. We consider a fixed set of $S = 9$ interval candidates with increasing interval lengths; and for all t, the lengths are $\{14, 30, 45, 60, 90, 120, 180, 240, 300\}$. The critical values are calibrated using the true underlying parameter, i.e., $\theta^* = \theta_0$, as well as the default interval candidates. In particular, 300 FAR curves are generated by using θ^*, where 300 corresponds to the longest interval length. The generation is repeated 1000 times and used to calibrate the critical values. The calibrated critical values are shown in Figure 4.3.1.

The estimation results of the misspecified AFAR model are listed in the last 3 columns of Table 4.1, including the average values of the estimators, as well as the estimation accuracy measured by RMSE and MAD. The AFAR model, though misspecified, provides reasonable estimation accuracy. The largest discrepancy from the true DGP case happens in estimating c_4, with the adaptive estimator

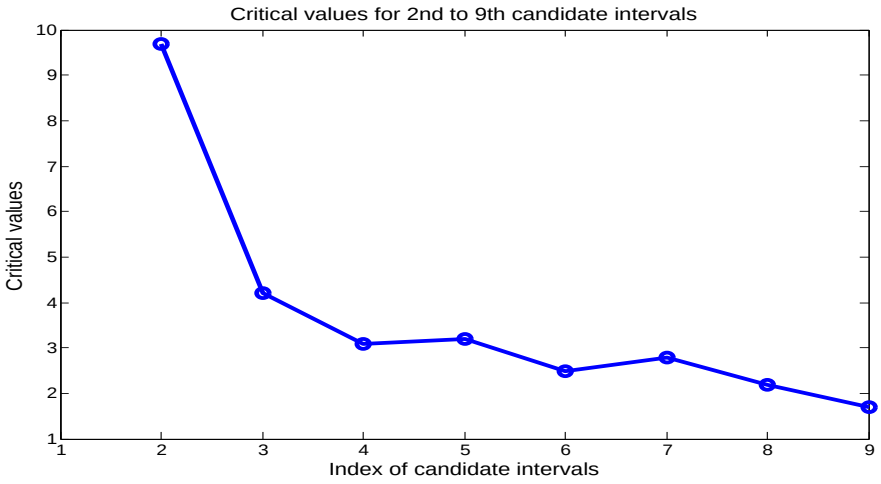

Figure 4.3.1 Critical values calculated for the second to the ninth candidate intervals.

being 0.96 instead of 1.01, the RMSE of 0.12 and the MAD of 0.07. We can see that the results are similar in both cases. In the HOM scenario, the theoretical selection of the local interval by AFAR should be the longest interval with $S = 9$, while in the simulation study, the average values of the selected local intervals are consistently around 7 or 8, which explains the comparable performance of AFAR model under stationarity.

4.3.2 Non-stationarity: Scenarios with regime shifts

Now we move to the scenarios with regime shifts and investigate the estimation accuracy of the AFAR modeling. Simple experiments are designed where only

one Fourier coefficient is changing at each time, and the rest remain constant. Based on the functional data generated in the HOM scenario, 10 experiments are considered, denoted as Scenario RS-X each, where the affixed notation X refers to the parameter that changes over time. In particular for each time, we generate 1500 functional data curves, with the first 500 curves (phase 1) from the underlying FAR model with θ_0, the next 500 curves (phase 2) from another FAR model with the labeled parameter shifting to a new level, and the last 500 curves (phase 3) again from the FAR model with the parameter setting back to the original level, see Table 4.2 under columns (in phase 2 and phase 3) entitled with "true" for the parameter changes in each scenario. In this way, the modeling structure always changes at $t = 501$ and 1001. The generation is also repeated for 1000 times.

For the non-stationary data, a direct application of the ML estimation is not appropriate any more. We instead employ the AFAR modeling. Compared to the rolling window technique, AFAR conducts estimation over selected local intervals, but with time-dependent interval lengths. Ideally, the selected local intervals should be long before the structural changes and short after the changes. As the interval candidates' lengths increase discontinuously from 14 ($s = 1$) to 300 ($s = 9$), the optimal selection would have a staircase pattern, lifting up when there are enough data for the longer interval. As an illustration, Figure 4.3.2 displays the average values of the selected intervals and the ideal interval selection for four RS

scenarios corresponding to four changing parameters. The adaptive approach selects reasonable local intervals, with large interval indexes before the first change, then shrinking to very small indexes after the change and gradually bouncing back. The time-dependent selection is advantageous for achieving information efficiency by selecting the long intervals for local stationarity and simultaneously quickly detecting changes of the instability.

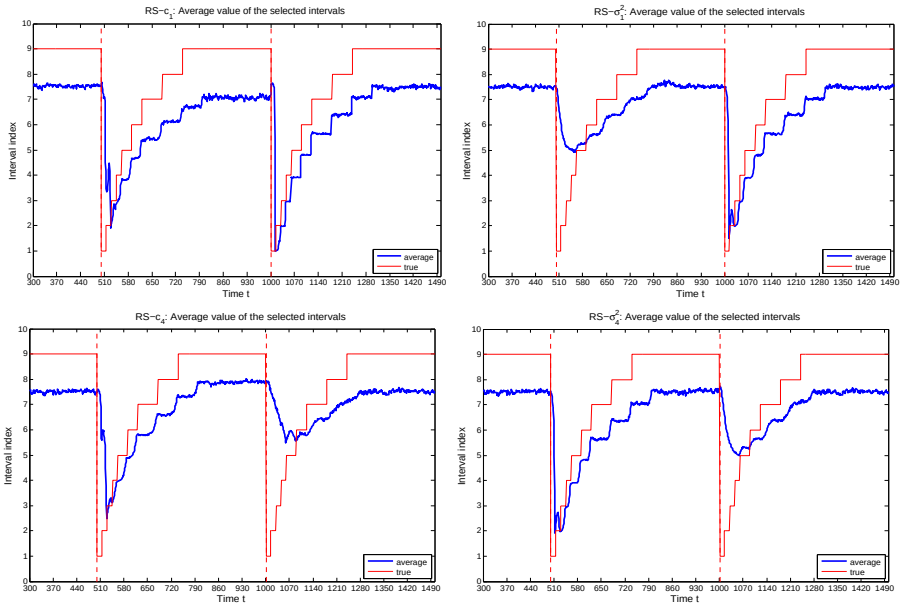

Figure 4.3.2 Detection accuracy: Average of the selected interval indexes from time 301 to 1500 in four RS scenarios, RS-c_1 (upper left), RS-c_4 (lower left), RS-σ_1^2 (upper right) and σ_4^2 (lower right), with only the affixed parameter changing over time. The blue curves indicate the trajectory of the average selected intervals and the red stepwise curves indicate the trajectory of the true theoretical intervals.

With the adaptive selection of local intervals, the AFAR modeling generally provides reasonable accuracy for non-stationary data. Table 4.2 reports the estimation results for all the RS scenarios with focus on the time-varying parameters in phase 2 and 3 only. Compared to the true parameters in each phase, the estimated coefficients are quite accurate, with RMSE smaller than 0.21 and MAD smaller than 0.15. The estimation of the other time-independent parameters does not deviate much from the homogeneous case. We hence only report the largest differences between the estimated parameters and those obtained in HOM scenario for the unchanged parameters. In the table, we denote the largest differences as LD, which in most cases are small. The detection delay is also calculated for each of the RS scenarios. Table 4.3 shows the average numbers of steps needed to reach 50%, 60%, 70% and 80% of the true parameter values for phase 2 and phase 3.

Besides, we investigate the estimation results when the parameter changes with different magnitudes and directions for scenario RS-c_1. The changes include upward large, upward small, downward large and downward small changes. Table 4.4 can be referred to for the four different changes for parameter c_1. In the table, the average values of the estimated coefficient, RMSE, MAD and LD for unchanged parameters are reported for both phase 2 and phase 3. Figure 4.3.3 displays the average selected intervals for the four cases. The average numbers of steps needed to reach a percentage of the true parameter values are calculated and shown in

Table 4.5.

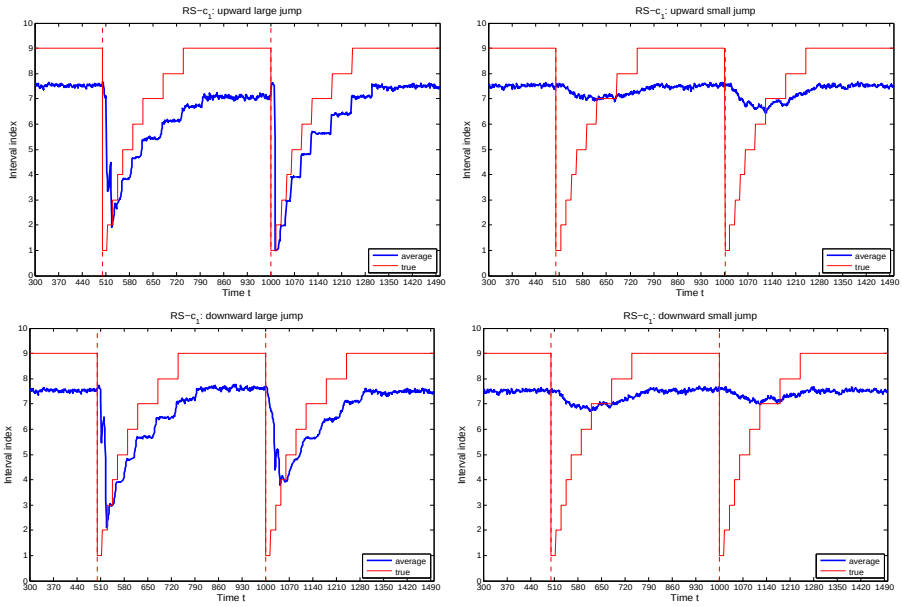

Figure 4.3.3 Detection accuracy: Average of the selected interval indexes from time 301 to 1500 in RS-c_1 scenario, with upward large, upward small, downward large and downward small jumps.

4.3.3 Robustness Checking

The estimation and forecast accuracy of the adaptive model depends on the critical values which themselves depend on the underlying hyperparameters, interval candidates and θ^*. As an illustration, we analyze the robustness of the hyperparameter choices under the RS-c_1 scenario, with the parameter c_1 being

time varying. The default values are $S = 9$ and $\theta^* = \theta_0$. We report the estimation accuracy for different sets of interval candidates. In particular, we consider using a set of intervals with fewer candidates, i.e., only the first 6 candidates with interval lengths $[14, 30, 45, 60, 90, 120]$, and a set of more candidates with interval lengths $[14, 30, 45, 60, 90, 120, 180, 240, 300, 360, 420, 500]$. The two cases are denoted by $S = 6$ and $S = 12$ respectively. Besides, we also use a sparse set and an intensive set of interval candidates. We have $S = 5$ with interval lengths $[14, 45, 90, 150, 300]$ for the sparse set and $S = 12$ with interval lengths $[14, 30, 45, 60, 90, 120, 150, 180, 210, 240, 270, 300]$ for the intensive set. Moreover, we analyze the impact on the estimation performance when θ^* is misspecified, with $\pm 20\%$, $\pm 30\%$, $\pm 40\%$ and $\pm 50\%$ deviation from c_1, its true value in θ_0. The misspecified cases are denoted as mis0.8 and mis1.2, mis0.7 and mis1.3, mis0.6 and mis1.4, mis0.5 and mis1.5. The other parameters still take the original true values. We also study four cases when the α value in Equation (4.10) deviates from 1. The four cases include $\alpha = 0.6$, $\alpha = 0.8$, $\alpha = 1.2$ and $\alpha = 1.4$.

In each of the four robustness experiments, critical values are re-calibrated, with which we perform the local interval selection and parameter estimation. The results are reported in Table 4.6. Compared to the default case with $S = 9$ and $\theta^* = \theta_0$, there is no big difference. To differentiate the numerical differences, we show the results with three decimal places. It suggests that, with an adaptive

and data-driven computation of the critical values, the estimation accuracy is, in general, robust with respect to the possible parameter misspecifications and different sets of interval candidates.

4.4. Real Data Analysis

In empirical analysis, we consider the hourly log-prices for California (CA) electricity market from 5 July 1999 to 11 June 2000, which have been studied in Weron and Misiorek (2008). After smoothing with B-splines, we obtain in total 343 log-price curves, see Figure 1.3.1. The first 300 curves corresponding to the dates 5 July 1999 – 29 April 2000 are used as a training set. We implement the AFAR modeling to forecast the electricity price curves from date 30 April 2000 onwards. The curves are expanded using 23 Fourier basis functions and the dimension of the sieve is 12. We adopt the same interval candidates used in the simulation study and use the ML estimators of the training set to generate Monte Carlo data, based on which the critical values are calibrated to achieve the best forecast accuracy. Starting from the 301st curve, i.e., the curve on 30 April 2000, and ending at the last curve on 11 June 2000, we apply the adaptive technique to identify an interval of local homogeneity at any time point in the period, over which the parameters are estimated. The fitted model is then used to iteratively compute one-day ahead

and two-week ahead forecasts of the electricity price curve.

For the purpose of comparison, we consider another three kinds of alternative models. They are constant FAR models, multivariate VAR models and univariate ARX, AR(1) and seasonal AR models. The constant FAR models are updated via the rolling window technique with a fixed window size of 300 and 150, denoted as FAR(300) and FAR(150). We take the length of the longest interval candidate, which is 300 for FAR(300) model and its half length, which is 150, for FAR(150) model. The multivariate VAR(1) model considers the cross-dependence among the 24 hourly log-price processes; while the univariate models are used to estimate and forecast the log-prices at each hour individually. In the three univariate models, some or all of the following predictors are included. The predictors are the lagged log-prices of the previous 1, 2 and 7 days, the minimum of the previous day's 24 hourly log-prices, the log-load forecast and three dummy variables corresponding to Mondays, Saturdays and Sundays for the weekly seasonality. These models are selected for their nice forecast performance in Weron and Misiorek (2008).

model	form	domain
AFAR	$X_t(\tau) - \mu_t(\tau) = \rho_t\big(X_{t-1}(\tau) - \mu_{t-1}(\tau)\big) + \varepsilon_t(\tau)$	$X_t(\tau), \quad \tau \in [0,1]$
FAR	$X_t(\tau) - \mu(\tau) = \rho\big(X_{t-1}(\tau) - \mu(\tau)\big) + \varepsilon_t(\tau)$	$X_t(\tau), \quad \tau \in [0,1]$
VAR	$X_t = (X_{1t}, \cdots, X_{dt})' = C + AX_{t-1} + \varepsilon_t$	$X_t \in \mathbb{R}^d, \quad C_{d\times 1}, A_{d\times d}$
AR(1)	$X_{jt} = \phi_{j0} + \phi_{j1}X_{j,t-1} + \varepsilon_{jt}$	$X_{jt} \in \mathbb{R}, \quad j = 1, \cdots, d$
ARseasonal	$X_{jt} = \phi_{j0} + \phi_{j1}X_{j,t-1} + \phi_{j2}X_{j,t-2} + \phi_{j3}X_{j,t-7} + \varepsilon_{jt}$	$X_{jt} \in \mathbb{R}, \quad j = 1, \cdots, d$
ARX	$X_{jt} = \phi_{j0} + \phi_{j1}X_{j,t-1} + \phi_{j2}X_{j,t-2} + \phi_{j3}X_{j,t-7}$ $+\theta_{j1}MinPrice_{t-1} + \theta_{j2}LoadForecast_{j,t-1}$ $+d_{j1}Mon + d_{j2}Sat + d_{j3}Sun + \varepsilon_{jt}$	$X_{jt} \in \mathbb{R}, \quad j = 1, \cdots, d$

Table 4.7 reports the RMSE values of the 43 one-day ahead out-of-sample forecasts for each of the 24 hours, with the best accuracy marked in bold. The results show that the AFAR model is the most successful, with 13 times out of 24 being the best forecast model among all the models. It is superior as it considers the cross-dependence of multiple stochastic processes. The AFAR model leads the best univariate model ARX, 15 times out of 24. Secondly, the FAR models, with parsimonious representation and hence fewer parameters, outperform its comparable alternative, the multivariate VAR model. The constant FAR(150) model beats VAR(1) model 12 times out of 24; the FAR(300) wins 17 times; and the AFAR model even leads 18 times. Moreover, the AFAR model, with adaptive estimation, is more accurate than the constant FAR models, with 19 out of 24 lower forecast errors. The median value of the selected interval lengths is 180. It reflects that the improved accuracy of AFAR is not due to the utilization of

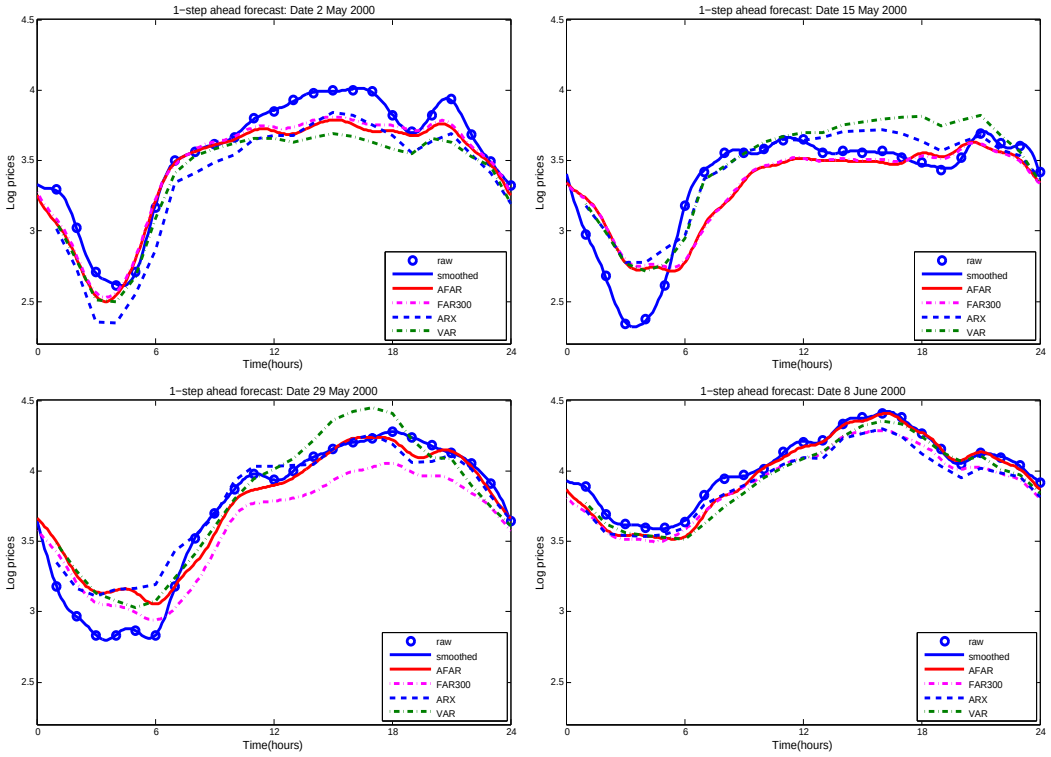

Figure 4.4.1 1-day ahead forecasted electricity log-price curves for dates 2 May, 15 May, 29 May and 8 June 2000 by AFAR, FAR(300), ARX and VAR(1).

short intervals such as the interval of length 150; instead, the adaptively selected intervals with time-varying lengths contribute significantly to the improvement of the forecast accuracy. It is advantageous to use AFAR model which considers the cross-dependence with parsimonious representation and simultaneously is flexible to provide reasonable forecast accuracy for both stationary and non-stationary functional data. As an illustration, Figure 4.4.1 shows the predicted curves for arbitrarily selected dates on 2 May, 15 May, 29 May and 8 June 2000 by AFAR,

FAR(300), ARX and VAR models. The shape and trend of the realized electricity price curves are well recovered by AFAR and FAR models. For ARX and VAR approaches, the price curves are obtained by drawing lines between each pair of adjacent data points. A longer forecast horizon of 14 days is also considered. The prediction results for 14-day ahead forecasts are shown in Table 4.8 and Figure 4.4.2. Similar prediction performance is obtained.

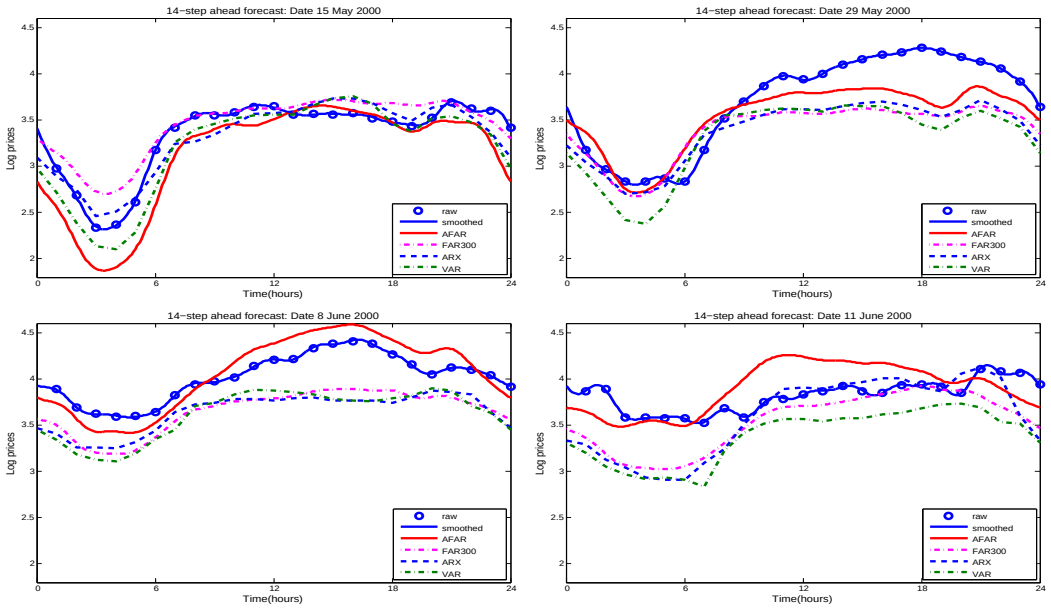

Figure 4.4.2 14-day ahead forecasted electricity log-price curves for dates 15 May, 29 May, 8 June and 11 June 2000 by AFAR, FAR(300), ARX and VAR(1).

		true DGP (FAR)			misspecified case (AFAR)		
	true	**coeff**	**RMSE**	**MAD**	**coeff**	**RMSE**	**MAD**
c_0	-0.78	-0.77	0.04	0.03	-0.76	0.07	0.05
c_1	0.39	0.38	0.06	0.05	0.37	0.10	0.07
c_2	-0.46	-0.46	0.05	0.04	-0.46	0.08	0.06
c_3	0.52	0.51	0.05	0.04	0.49	0.10	0.06
c_4	1.01	1.00	0.04	0.03	0.96	0.12	0.07
σ_0^2	0.14	0.14	0.01	0.01	0.14	0.02	0.01
σ_1^2	0.03	0.02	0.00	0.00	0.02	0.00	0.00
σ_2^2	0.10	0.09	0.01	0.00	0.09	0.01	0.01
σ_3^2	0.18	0.18	0.01	0.01	0.18	0.02	0.01
σ_4^2	0.10	0.10	0.01	0.00	0.10	0.01	0.01
p_0	0.31	0.31	0.02	0.02	0.31	0.03	0.02
q_1	0.37	0.37	0.02	0.02	0.38	0.04	0.03
p_1	0.43	0.44	0.03	0.02	0.44	0.04	0.03
q_2	0.18	0.18	0.02	0.01	0.18	0.03	0.02
p_2	0.36	0.36	0.02	0.02	0.36	0.03	0.02
q_3	0.50	0.51	0.04	0.03	0.52	0.07	0.04
p_3	-0.24	-0.24	0.03	0.02	-0.25	0.05	0.03
q_4	0.20	0.21	0.03	0.02	0.23	0.07	0.04
p_4	-0.18	-0.19	0.03	0.02	-0.20	0.07	0.04

Table 4.1 Finite sample estimation accuracy for scenario HOM. The misspecified estimation with AFAR modeling is compared with the true data generating process (DGP) of FAR modeling.

	phase 2: time index from 501 to 1000					phase 3: time index from 1001 to 1500				
	true	**coeff**	**RMSE**	**MAD**	**LD**	**true**	**coeff**	**RMSE**	**MAD**	**LD**
RS-c_0	0.78	0.67	0.21	0.13	0.10	-0.78	-0.72	0.13	0.09	0.02
RS-c_1	1.10	1.02	0.17	0.10	0.08	0.39	0.38	0.13	0.10	0.01
RS-c_2	-1.20	-1.13	0.12	0.10	0.01	-0.46	-0.52	0.15	0.12	0.01
RS-c_3	-0.20	-0.16	0.14	0.12	0.01	0.52	0.42	0.16	0.13	0.00
RS-c_4	0.10	0.13	0.14	0.12	0.02	1.01	0.90	0.18	0.13	0.01
RS-σ_0^2	0.01	0.01	0.01	0.00	0.02	0.14	0.12	0.03	0.03	0.01
RS-σ_1^2	0.30	0.27	0.05	0.05	0.01	0.03	0.03	0.01	0.01	0.02
RS-σ_2^2	1.00	0.89	0.18	0.15	0.01	0.10	0.12	0.04	0.03	0.02
RS-σ_3^2	0.01	0.01	0.01	0.00	0.02	0.18	0.16	0.03	0.03	0.01
RS-σ_4^2	0.01	0.01	0.00	0.00	0.01	0.10	0.09	0.02	0.02	0.01

Table 4.2 RS scenario: estimation of the parameters when there is a sudden change for one of the parameters. Each row reports the estimation results for the changed parameter only. The second to the fifth columns contain the average values of the estimated parameters, RMSE, MAD of the estimators and the largest deviation (LD) of the estimates for those unchanged parameters from the HOM scenario for phase 2. The last five columns contain results for phase 3.

	phase 2				phase 3			
	50%	**60%**	**70%**	**80%**	**50%**	**60%**	**70%**	**80%**
RS-c_0	24	25	27	29	13	13	13	13
RS-c_1	4	5	6	7	13	13	13	13
RS-c_2	12	23	36	51	43	44	56	58
RS-c_3	56	57	57	58	55	60	79	89
RS-c_4	29	44	44	44	27	35	44	57
RS-σ_0^2	29	29	29	29	49	60	76	89
RS-σ_1^2	39	49	57	81	13	28	28	28
RS-σ_2^2	39	47	57	79	28	28	28	28
RS-σ_3^2	13	13	13	13	39	46	54	77
RS-σ_4^2	28	28	28	28	40	48	57	79

Table 4.3 Detection delay for RS scenarios: the first four columns contain the average number of steps needed to reach 50%, 60%, 70% and 80% of the true values for phase 2. The last four columns contain results for phase 3.

	phase 2: time index from 501 to 1000					phase 3: time index from 1001 to 1500				
	true	**coeff**	**RMSE**	**MAD**	**LD**	**true**	**coeff**	**RMSE**	**MAD**	**LD**
up large	1.10	1.02	0.17	0.10	0.08	0.39	0.38	0.13	0.10	0.01
up small	0.51	0.47	0.10	0.07	0.01	0.39	0.39	0.11	0.08	0.01
down large	-0.32	-0.29	0.13	0.11	0.01	0.39	0.34	0.13	0.10	0.01
down small	0.27	0.28	0.10	0.08	0.00	0.39	0.35	0.10	0.07	0.00

Table 4.4 RS-c_1 scenario with upward large, upward small, downward large and downward small jumps: Each row reports the estimation results for the changed parameter only. The second to the fifth columns contain the average values of the estimated parameters, RMSE, MAD of the estimators and the largest deviation (LD) of the estimates for those unchanged parameters from the HOM scenario for phase 2. The last five columns contain results for phase 3.

	phase 2				phase 3			
	50%	**60%**	**70%**	**80%**	**50%**	**60%**	**70%**	**80%**
up large	4	5	6	7	13	13	13	13
up small	1	1	1	24	1	1	1	31
down large	28	29	29	29	26	29	34	37
down small	1	1	33	57	1	1	14	58

Table 4.5 Detection delay for RS-c_1 scenario with upward large, upward small, downward large and downward small changes. The first four columns contain the average number of steps needed to reach 50%, 60%, 70% and 80% of the true values for phase 2. The last four columns contain results for phase 3.

	phase 2: time index from 501 to 1000					phase 3: time index from 1001 to 1500				
	true	**coeff**	**RMSE**	**MAD**	**LD**	**true**	**coeff**	**RMSE**	**MAD**	**LD**
default	1.100	1.024	0.165	0.099	0.078	0.390	0.378	0.133	0.102	0.013
mis0.8	1.100	1.026	0.165	0.099	0.078	0.390	0.378	0.132	0.102	0.012
mis1.2	1.100	1.022	0.170	0.102	0.083	0.390	0.377	0.135	0.103	0.015
mis0.7	1.100	1.027	0.164	0.098	0.076	0.39	0.378	0.132	0.102	0.010
mis1.3	1.100	1.020	0.175	0.103	0.086	0.390	0.380	0.135	0.103	0.013
mis0.6	1.100	1.020	0.172	0.103	0.086	0.390	0.377	0.136	0.104	0.015
mis1.4	1.100	1.020	0.171	0.103	0.085	0.390	0.376	0.134	0.103	0.015
mis0.5	1.100	1.019	0.175	0.103	0.087	0.390	0.376	0.137	0.105	0.015
mis1.5	1.100	1.027	0.159	0.098	0.076	0.390	0.378	0.133	0.103	0.013
S=6	1.100	0.982	0.211	0.139	0.137	0.390	0.359	0.170	0.132	0.044
S=12	1.100	1.033	0.153	0.091	0.065	0.390	0.382	0.124	0.095	0.005
sparse	1.100	1.006	0.197	0.117	0.106	0.390	0.371	0.147	0.112	0.021
intensive	1.100	1.031	0.156	0.093	0.069	0.390	0.380	0.127	0.099	0.008
$\alpha = 0.8$	1.100	1.032	0.157	0.093	0.068	0.390	0.381	0.127	0.098	0.008
$\alpha = 1.2$	1.100	1.021	0.168	0.102	0.084	0.390	0.377	0.135	0.104	0.015
$\alpha = 0.6$	1.100	1.039	0.146	0.086	0.058	0.390	0.386	0.122	0.096	0.006
$\alpha = 1.4$	1.100	1.017	0.175	0.105	0.090	0.390	0.375	0.140	0.108	0.018

Table 4.6 Robustness checking in RS-c_1 scenario: "mis0.8" and "mis1.2", "mis0.7" and "mis1.3" , "mis0.6" and "mis1.4" , "mis0.5" and "mis1.5" refer to the misspecified cases where the underlying parameter is biased with ±20%, ±30%, ±40% and ±50% deviation. "$S = 6$" and "$S = 12$" refer to the cases with fewer and more interval candidates. "sparse" and "intensive" refer to a sparse set with 5 interval candidates and an intensive set with 12 candidates. Four cases for different α values are also studied.

Hours	AFAR	FAR (300)	FAR (150)	VAR(1)	ARX	AR(1)	ARseasonal
1	**0.167**	0.170	0.184	0.171	0.178	0.185	0.182
2	**0.222**	0.223	0.234	0.225	0.227	0.232	0.234
3	0.271	0.265	0.273	**0.260**	0.269	0.274	0.282
4	0.256	0.250	0.259	**0.236**	0.246	0.254	0.261
5	0.243	0.239	0.249	**0.230**	0.239	0.254	0.260
6	0.283	0.273	0.279	0.271	**0.236**	0.274	0.271
7	0.271	0.267	0.273	0.255	**0.179**	0.269	0.207
8	0.231	0.234	0.243	0.225	**0.163**	0.241	0.193
9	0.169	0.178	0.192	0.177	**0.151**	0.185	0.171
10	**0.203**	0.229	0.248	0.241	0.224	0.237	0.237
11	**0.263**	0.287	0.307	0.311	0.291	0.298	0.305
12	**0.276**	0.304	0.324	0.326	0.297	0.310	0.309
13	**0.297**	0.322	0.343	0.357	0.320	0.332	0.336
14	**0.325**	0.349	0.367	0.389	0.335	0.352	0.357
15	0.338	0.355	0.377	0.396	**0.326**	0.355	0.351
16	0.356	0.368	0.391	0.401	**0.338**	0.360	0.362
17	0.351	0.358	0.382	0.392	**0.334**	0.352	0.356
18	0.326	0.328	0.351	0.362	**0.312**	0.331	0.339
19	**0.303**	0.313	0.335	0.364	0.317	0.317	0.340
20	**0.290**	0.302	0.324	0.357	0.315	0.303	0.331
21	**0.280**	0.300	0.320	0.327	0.299	0.302	0.304
22	**0.255**	0.279	0.295	0.312	0.301	0.299	0.301
23	**0.131**	0.152	0.171	0.152	0.141	0.147	0.133
24	**0.117**	0.127	0.139	0.135	0.129	0.122	0.118

Table 4.7 1-day ahead forecasts: RMSE of the out-of-sample forecasts using the FAR models, VAR(1) model and univariate models. In particular, the AFAR forecasts are compared with the FAR updated with rolling window technique of fixed window size 150 and 300, VAR(1), ARX, AR(1) and seasonal AR models.

Hours	AFAR	FAR (300)	FAR (150)	VAR(1)	ARX	AR(1)	ARseasonal
1	0.561	**0.397**	0.491	0.562	0.474	0.418	0.420
2	0.673	**0.416**	0.554	0.636	0.449	0.423	0.508
3	0.797	**0.463**	0.664	0.698	0.521	0.484	0.630
4	0.800	**0.470**	0.690	0.715	0.511	0.479	0.629
5	0.758	**0.415**	0.621	0.643	0.423	0.427	0.549
6	0.612	**0.308**	0.458	0.500	0.349	0.326	0.416
7	0.436	**0.293**	0.371	0.475	0.347	0.297	0.317
8	0.381	0.307	0.373	0.398	0.356	0.303	**0.286**
9	0.353	0.343	0.392	0.390	0.380	0.348	**0.324**
10	**0.441**	0.479	0.504	0.491	0.480	0.482	0.457
11	**0.519**	0.586	0.602	0.561	0.564	0.587	0.559
12	**0.533**	0.629	0.645	0.600	0.607	0.630	0.601
13	**0.575**	0.674	0.695	0.644	0.651	0.679	0.654
14	**0.603**	0.722	0.745	0.685	0.695	0.731	0.697
15	**0.629**	0.740	0.770	0.724	0.712	0.753	0.732
16	**0.649**	0.769	0.803	0.753	0.738	0.779	0.762
17	**0.650**	0.766	0.815	0.759	0.737	0.770	0.771
18	**0.623**	0.703	0.764	0.734	0.682	0.694	0.742
19	**0.608**	0.647	0.701	0.668	0.633	0.640	0.701
20	**0.530**	0.596	0.637	0.586	0.575	0.597	0.613
21	**0.526**	0.607	0.643	0.597	0.574	0.605	0.569
22	**0.518**	0.579	0.597	0.606	0.561	0.591	0.552
23	0.394	0.411	0.463	0.465	0.436	0.442	**0.366**
24	0.483	**0.388**	0.464	0.538	0.470	0.423	0.394

Table 4.8 14-day ahead forecasts: RMSE of the out-of-sample forecasts using the FAR models, VAR(1) model and univariate models. In particular, the AFAR forecasts are compared with the FAR updated with rolling window technique of fixed window size 150 and 300, VAR(1), ARX, AR(1) and seasonal AR models.

CHAPTER 5

Conclusion and future work

In Chapter 2, we note that Diebold and Li (2006) propose the DNS model, which shows nice performance in forecasting U.S. yield curves. Nevertheless, it depends on the NS factor loadings, which are fixed and may misrepresent the underlying structure of yield curves. It is also observed that the NS factors are persistent, and the stationary AR(1) process is unable to replicate the factors. In this thesis, we propose the FPCA-LAR model, which is data-driven and able to account for non-stationarity in yield curves. On the basis of the yield curves in U.S. and China, simulation study and real data analysis illustrate that the proposed method performs well in comparison to the DNS model. In particular, the proposed

FPCA-LAR model improves the 6-month and 12-month ahead forecasts. The analysis also reveals that FPCA-LAR is a flexible and adaptive approach capable of capturing the underlying structure for any type of yield curves in different markets. Especially, for the data with data structure deviating from the exponential basis functions, the proposed method makes a better choice for forecasting. In this thesis, we mainly focus on forecasting yield curves within one data sample, but the cross dependence between U.S. and China markets is also of great interest to study. It is worthwhile to mention that Benko, Härdle and Kneip (2009) have proposed Common FPCA for two or more than two samples inference, which makes an essential reference for the future research of U.S. and China markets.

In Chapter 3, we extend the univariate adaptive LAR model to multivariate LVAR model which is capable of providing reasonable forecast accuracy under both homogeneity and structural changes. Compared to the conventional dynamic time series models using either a rolling window with a globally fixed window size or a recursive technique using all past information, the adaptive procedure carefully selects an interval of local homogeneity at any particular time point. With this flexibility on interval selection, the LVAR model provides stable performance both in a simulated homogeneous situation and under regime shift scenarios. The real data analysis provides an example of real-time monitoring and forecasting of the yield curves. The selected intervals indicate that there are some underlying

economic interpretations behind the detection results, thus LVAR is useful for monitoring purposes. Although the forecast results are not satisfying, which may be due to frequent changes that are not easily detected by a high-dimensional model setting without penalizing parameter uncertainty, we note from existing works that imposing specification structures could improve predictability. We relegate the discussion of further improvement within the general LVAR framework to future work.

In Chapter 4, we propose AFAR modeling for estimating and forecasting both stationary and non-stationary functional autoregressive data. Under stationarity, where the AFAR model drops to the constant FAR model, we develop consistent ML estimators with closed forms. The AFAR modeling is flexible, which incorporates various sources of changes without explicitly assuming change types and timing. Under non-stationarity, we select the interval of local homogeneity and safely conduct the ML estimation over the local interval. The consistency result of the sieve ML estimator and the theoretical properties of the adaptive functional estimate are studied and proved. Simulation study shows that the AFAR model provides comparable accuracy with the true data generating process under a homogeneous scenario and reasonable performance when there are regime shifts. Moreover, the adaptive estimation is robust to the selection of hyperparameters

that are used to calibrate critical values. In real data analysis on forecasting log-prices of California electricity market, the AFAR modeling is the most successful, compared to the alternatives including the constant FAR models, VAR model and several univariate models. For the future work, exogenous variables and variables taking care of the seasonal pattern of electricity prices will be introduced into the model. The estimation of parameters corresponding to exogenous variables are considered to be similar as what we have done for estimating the mean function.

Appendix

Appendix A: Derivation of the Fourier coefficient relationship as shown in Section 4.1.1

Trigonometric basis functions in $L^2([0,1])$, that are $\phi_0 = 1_{[0,1]}$, $\phi_{2k}(\tau) = \sqrt{2}\cos 2\pi k\tau, \phi_{2k-1}(\tau) = \sqrt{2}\sin 2\pi k\tau$ are considered to expand all the involved processes. Therefore, the curves can be represented by Fourier basis and Fourier coefficients as

$$X_t(\tau) = a_{t,0} + \sum_{k=1}^{m_n} b_{t,k}\phi_{2k-1}(\tau) + a_{t,k}\phi_{2k}(\tau),$$

$$\varepsilon_t(\tau) = a_0(\varepsilon_t) + \sum_{k=1}^{m_n} b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau)$$

and

$$K(\tau) = c_0 + \sum_{k=1}^{m_n} d_k\phi_{2k-1}(\tau) + c_k\phi_{2k}(\tau),$$

where $a_{t,0}, a_{t,k}, b_{t,k}$ are the Fourier coefficients corresponding to the constant, cosine, and sine basis functions respectively for curve X_t; $a_0(\varepsilon_t), a_k(\varepsilon_t), b_k(\varepsilon_t)$ are the Fourier coefficients corresponding to the constant, cosine, and sine basis functions respectively for curve ε_t; c_0, c_k, d_k are the Fourier coefficients corresponding to the constant, cosine, and sine basis functions respectively for curve K.

When the mean function $\mu(\tau)$ is a zero constant function, the Equation (4.2) can be rewritten as

$$X_t(\tau) = a_{t,0} + \sum_{k=1}^{m_n} b_{t,k}\phi_{2k-1}(\tau) + a_{t,k}\phi_{2k}(\tau)$$

$$= \int_0^1 [c_0 + \sum c_k\phi_{2k}(\tau - s) + \sum d_k\phi_{2k-1}(\tau - s)][a_{t-1,0} + \sum a_{t-1,k}\phi_{2k}(s)$$

$$+ \sum b_{t-1,k}\phi_{2k-1}(s)]ds + a_0(\varepsilon_t) + \sum_{k=1}^{m_n} b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau)$$

By the properties of Fourier basis functions, we have

$$\int_0^1 \phi_i(\tau)\phi_j(\tau)d\tau = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$$\int_0^1 \phi_k(\tau)d\tau = 0,$$

$$\int_0^1 \phi_{2i}(\tau - s)\phi_{2j}(s)ds = \begin{cases} 0 & \text{if } i \neq j \\ \frac{\phi_{2i}(\tau)}{\sqrt{2}} & \text{if } i = j \end{cases}$$

$$\int_0^1 \phi_{2i}(\tau - s)\phi_{2j-1}(s)ds = \begin{cases} 0 & \text{if } i \neq j \\ \frac{\phi_{2i-1}(\tau)}{\sqrt{2}} & \text{if } i = j \end{cases}$$

$$\int_0^1 \phi_{2i-1}(\tau-s)\phi_{2j}(s)ds = \begin{cases} 0 & \text{if } i \neq j \\ \frac{\phi_{2i-1}(\tau)}{\sqrt{2}} & \text{if } i = j \end{cases}$$

$$\int_0^1 \phi_{2i-1}(\tau-s)\phi_{2j-1}(s)ds = \begin{cases} 0 & \text{if } i \neq j \\ -\frac{\phi_{2i}(\tau)}{\sqrt{2}} & \text{if } i = j \end{cases}$$

Therefore, we have

$$X_t(\tau) = a_{t,0} + \sum_{k=1}^{m_n} b_{t,k}\phi_{2k-1}(\tau) + a_{t,k}\phi_{2k}(\tau)$$

$$= c_0 a_{t-1,0} + \sum \frac{1}{\sqrt{2}} c_k a_{t-1,k}\phi_{2k}(\tau) + \sum \frac{1}{\sqrt{2}} c_k b_{t-1,k}\phi_{2k-1}(\tau)$$
$$+ \sum \frac{1}{\sqrt{2}} d_k a_{t-1,k}\phi_{2k-1}(\tau) - \sum \frac{1}{\sqrt{2}} d_k b_{t-1,k}\phi_{2k}(\tau)$$
$$+ a_0(\varepsilon_t) + \sum b_k(\varepsilon_t)\phi_{2k-1}(\tau) + a_k(\varepsilon_t)\phi_{2k}(\tau).$$

Hence the coefficient relationship satisfies

$$a_{t,0} = c_0 a_{t-1,0} + a_0(\varepsilon_t)$$

$$a_{t,k} = \frac{1}{\sqrt{2}} c_k a_{t-1,k} - \frac{1}{\sqrt{2}} d_k b_{t-1,k} + a_k(\varepsilon_t)$$
$$b_{t,k} = \frac{1}{\sqrt{2}} c_k b_{t-1,k} + \frac{1}{\sqrt{2}} d_k a_{t-1,k} + b_k(\varepsilon_t)$$

In the paper, we assume that the kernel function K is an even function, that is all the $d_k = 0$.

When the mean function is non-zero, we have

$$X_t(\tau) = \mu(\tau) - \int_0^1 K(\tau-s)\mu(s)ds + \int_0^1 K(\tau-s)X_{t-1}(s)ds + \varepsilon_t(\tau)$$

If we assign p_0, p_k, q_k being the Fourier coefficients corresponding to the constant, consine and sine basis functions respectively for function $\mu(\tau) - \int_0^1 K(\tau - s)\mu(s)ds$, we will have

$$\begin{aligned} a_{t,0} &= p_0 + c_0 a_{t-1,0} + a_0(\varepsilon_t) \\ a_{t,k} &= p_k + \frac{1}{\sqrt{2}} c_k a_{t-1,k} - \frac{1}{\sqrt{2}} d_k b_{t-1,k} + a_k(\varepsilon_t) \\ b_{t,k} &= q_k + \frac{1}{\sqrt{2}} c_k b_{t-1,k} + \frac{1}{\sqrt{2}} d_k a_{t-1,k} + b_k(\varepsilon_t). \end{aligned}$$

This is exactly what we have utilized in developing AFAR when d_k is assumed to be zero in Section 4.1.1.

Appendix B: Proof of the consistency result in Theorem 4.1.2

Proof. Without loss of generality, we first study the case when p_0, p_k, q_k are all zeros. The same consistency results can be obtained for non-zero cases. We check condition C1. From now on, we replace m_n by m for notational simplicity. From the definition of the entropy, we have

$$\begin{aligned} &H(\rho_{0|\Theta_m}, \rho_{0|\Theta_m}) - H(\rho_{0|\Theta_m}, \rho_m) = H(K_{0|\Theta_m}, K_{0|\Theta_m}) - H(K_{0|\Theta_m}, K_m) \\ &= -\log\sigma_0 + \log\sigma_{0,m} - \sum_{i=1}^{m} \log(\sigma_i^2) + \sum_{i=1}^{m} \log(\sigma_{i,m}^2) \\ &+ E\Big(- \frac{1}{2\sigma_0^2}(a_{t,0} - c_0 a_{t-1,0})^2 + \frac{1}{2\sigma_{0,m}^2}(a_{t,0} - c_{0,m} a_{t-1,0})^2 \\ &- \sum_{k=1}^{m} \frac{1}{2\sigma_k^2}[(b_{t,k} - \frac{1}{\sqrt{2}} c_k b_{t-1,k})^2 + (a_{t,k} - \frac{1}{\sqrt{2}} c_k a_{t-1,k})^2] \\ &+ \sum_{k=1}^{m} \frac{1}{2\sigma_{k,m}^2}[(b_{t,k} - \frac{1}{\sqrt{2}} c_{k,m} b_{t-1,k})^2 + (a_{t,k} - \frac{1}{\sqrt{2}} c_{k,m} a_{t-1,k})^2]\Big), \end{aligned}$$

where $\sigma_0^2, \sigma_i^2, c_0$ and c_k denote the variances and Fourier coefficients for kernel $K_{0|\Theta_m}$ and $\sigma_{0,m}^2, \sigma_{i,m}^2, c_{0,m}$ and $c_{k,m}$ are for those of K_m. K_m is the kernel function for ρ_m with $\rho_m \in \Theta_m$; and $K_{0|\Theta_m}$ is the projection of the true kernel function K_0 on Θ_m. We assume that $\sigma_0 = \sigma_{0,m}$ and $\sigma_i = \sigma_{i,m}$, so we have

$$
\begin{aligned}
&H(\rho_{0|\Theta_m}, \rho_{0|\Theta_m}) - H(\rho_{0|\Theta_m}, \rho_m) \\
&= \frac{1}{2\sigma_0^2} E[-(a_{t,0} - c_0 a_{t-1,0})^2 + (a_{t,0} - c_{0,m} a_{t-1,0})^2] \\
&+ \sum_{k=1}^{m} \frac{1}{2\sigma_k^2} E[-(b_{t,k} - \frac{1}{\sqrt{2}} c_k b_{t-1,k})^2 - (a_{t,k} - \frac{1}{\sqrt{2}} c_k a_{t-1,k})^2 \\
&+ (b_{t,k} - \frac{1}{\sqrt{2}} c_{k,m} b_{t-1,k})^2 + (a_{t,k} - \frac{1}{\sqrt{2}} c_{k,m} a_{t-1,k})^2] \\
&= \frac{1}{2\sigma_0^2} E[(c_{0,m} a_{t-1,0} - c_0 a_{t-1,0})^2] \\
&+ \sum_{k=1}^{m} \frac{1}{2\sigma_k^2} E[\frac{1}{2}(c_{k,m} a_{t-1,k} - c_k a_{t-1,k})^2 + \frac{1}{2}(c_{k,m} b_{t-1,k} - c_k b_{t-1,k})^2].
\end{aligned}
$$

In the last step, we have

$$
\begin{aligned}
&E[-(a_{t,0} - c_0 a_{t-1,0})^2 + (a_{t,0} - c_{0,m} a_{t-1,0})^2] \\
&= E\big[E[-(a_{t,0} - c_0 a_{t-1,0})^2 + (a_{t,0} - c_{0,m} a_{t-1,0})^2 | a_{t-1,0}]\big] \\
&= E\big[E[(c_{0,m} a_{t-1,0} - c_0 a_{t-1,0})^2 + 2(c_0 a_{t-1,0} - c_{0,m} a_{t-1,0})(a_{t,0} - c_0 a_{t-1,0}) | a_{t-1,0}]\big] \\
&= E[(c_{0,m} a_{t-1,0} - c_0 a_{t-1,0})^2].
\end{aligned}
$$

The derivation is similar for other square terms. Therefore, if $H(K_{0|\Theta_m}, K_m) \to H(K_{0|\Theta_m}, K_{0|\Theta_m})$ as $n, m \to \infty$, we have $c_{0,m} \to c_0$, $c_{k,m} \to c_k$ and consequently $\rho_m \to \rho_{0|\Theta_m}$.

For the condition C2 and (i) of Theorem 4.1.1, similar arguments could be followed as in Mourid and Bensmain (2006). To check Theorem 4.1.1 (ii), we define

$$
\varphi(t) = E_{K_{0|\Theta_m}}\big(\exp[t \log \frac{g(X_t, X_{t-1}, \Gamma_k)}{g(X_t, X_{t-1}, K_m)}]\big),
$$

where $g(X_t, X_{t-1}, \Gamma_k) = \sup_{\beta\in\Gamma_k} g(X_t, X_{t-1}, \beta)$ and $g(X_t, X_{t-1}, K_m) = g(X_t, X_{t-1}, \rho_m)$. Furthermore, we have $\varphi(0) = 1$ and $\varphi'(0) = E_{K_{0|\Theta_m}} \log \frac{g(X_t, X_{t-1}, \Gamma_k)}{g(X_t, X_{t-1}, K_m)}$.

For a fixed $K \in \Gamma_k$, we have

$$\begin{aligned}
A &= E_{K_{0|\Theta_m}} \log g(X_t, X_{t-1}, \Gamma_k) - E\log g((X_t, X_{t-1}, K)) \\
&= E_{K_{0|\Theta_m}} \sup_{\beta\in\Gamma_k} \big[\log g(X_t, X_{t-1}, \beta) - \log g(X_t, X_{t-1}, K)\big] \\
&= E_{K_{0|\Theta_m}} \sup_{\beta\in\Gamma_k} \big[-\log\sigma_{0,\beta} + \log\sigma_{0,K} - \sum_{i=1}^{m} \log\sigma_{i,\beta}^2 + \sum_{i=1}^{m} \log\sigma_{i,K}^2 \\
&- \frac{1}{2\sigma_{0,\beta}^2}(a_{t,0} - c_{0,\beta}a_{t-1,0})^2 + \frac{1}{2\sigma_{0,K}^2}(a_{t,0} - c_{0,K}a_{t-1,0})^2 \\
&- \sum_{k=1}^{m} \frac{1}{2\sigma_{k,\beta}^2}[(b_{t,k} - \frac{1}{\sqrt{2}}c_{k,\beta}b_{t-1,k})^2 + (a_{t,k} - \frac{1}{\sqrt{2}}c_{k,\beta}a_{t-1,k})^2] \\
&+ \sum_{k=1}^{m} \frac{1}{2\sigma_{k,K}^2}[(b_{t,k} - \frac{1}{\sqrt{2}}c_{k,K}b_{t-1,k})^2 + (a_{t,k} - \frac{1}{\sqrt{2}}c_{k,K}a_{t-1,k})^2].
\end{aligned}$$

By assuming $\sigma_{k,\beta} = \sigma_{k,K} = \sigma_k, k = 0, ..., m$, we further have

$$\begin{aligned}
A = E_{K_{0|\Theta_m}} \sup_{\beta\in\Gamma_k} \big[&\frac{1}{2\sigma_0^2}(c_{0,K}a_{t-1,0} - c_{0,\beta}a_{t-1,0})(c_{0,K}a_{t-1,0} + c_{0,\beta}a_{t-1,0} - 2a_{t,0}) \\
&+ \sum_{k=1}^{m} \frac{1}{2\sigma_k^2}[(\frac{1}{\sqrt{2}}c_{k,K}b_{t-1,k} - \frac{1}{\sqrt{2}}c_{k,\beta}b_{t-1,k})(\frac{1}{\sqrt{2}}c_{k,\beta}b_{t-1,k} + \frac{1}{\sqrt{2}}c_{k,K}b_{t-1,k} - 2b_{t,k}) \\
&+ (\frac{1}{\sqrt{2}}c_{k,K}a_{t-1,k} - \frac{1}{\sqrt{2}}c_{k,\beta}a_{t-1,k})(\frac{1}{\sqrt{2}}c_{k,\beta}a_{t-1,k} + \frac{1}{\sqrt{2}}c_{k,K}a_{t-1,k} - 2a_{t,k})]\big].
\end{aligned}$$

By following the similar conditions and arguments in Mourid and Bensmain (2006), we can obtain $A \le \frac{C_1}{m^{\eta/2}}$, where C_1 is a constant. In addition, for $\delta > 0$,

$$\varphi'(0) = H(K_{0|\Theta_m}, K) - H(K_{0|\Theta_m}, K_m) + A \le C_2 m^{(-\eta/2)} - \delta.$$

By using Taylor expansion and the results from Hwang (1980) such that $\varphi''(t) \le C_3 m^2$, we have $\varphi(\frac{1}{m^2}) \le 1 - \frac{\delta}{C_4 m^2}$, where C_2, C_3 and C_4 are constants. Since $\varphi_m = \sup_k \inf_{t\ge 0} \varphi(t)$, we can deduce for m being sufficiently large, we have

$$l_m(\varphi_m)^n \le C m^{C m^{1+\eta}} (1 - \frac{\delta}{C m^2})^n,$$

which is summable if $m = \mathrm{O}(n^{1/3-\delta})$ for $\delta > 0$ (see Hwang (1980)). Finally, we can apply Theorem 4.1.1 to obtain the result that the ML estimator $\hat{K}$ obtained on Θ_{m_n} converges to the projected true kernel function $K_{0|\Theta_{m_n}}$. As $n, m_n \to \infty$, $K_{0|\Theta_{m_n}} \to K_0$ because $K_{0|\Theta_{m_n}}$ is just the Fourier truncation of true kernel K_0 on Θ_{m_n}. □

Appendix C: Proof of Theorem 4.2.2

Proof. The proof is based on the following general result.

Lemma. *Let P and P_0 be two measures such that the Kullback-Leibler divergence $E[\log(dP/dP_0)]$ satisfies*

$$E[\log(dP/dP_0)] \leq \Delta < \infty.$$

Then for any random variable ζ with $E_0\zeta < \infty$, it holds that

$$E[\log(1+\zeta)] \leq \Delta + E_0\zeta.$$

Proof. We can check that, for any fixed y, the maximum of the function $f(x) = xy - x\log x + x$ is attained at $x = e^y$, leading to the inequality $xy \leq x\log x - x + e^y$. Using this inequality and the equation $E[\log(1+\zeta)] = E_0[Z\log(1+\zeta)]$ with $Z = dP/dP_0$, we obtain

$$\begin{aligned} E[\log(1+\zeta)] &= E_0[Z\log(1+\zeta)] \\ &\leq E_0(Z\log Z - Z) + E_0(1+\zeta) \\ &= E_0(Z\log Z) + E_0\zeta - E_0 Z + 1. \end{aligned}$$

We note that $E_0(Z \log Z) = E(\log Z) = E\left[\log(d\mathbb{P}/d\mathbb{P}_0)\right] \leq \Delta$ and $E_0 Z = E_0(dP/dP_0) = 1$. □

We now apply this lemma with $\zeta = \rho(\hat{\theta}, \theta)/\mathfrak{R}(\hat{\theta}, \theta)$ and utilize the equation $E_0\zeta = E_\theta\left[\rho(\hat{\theta}, \theta)/\mathfrak{R}(\hat{\theta}, \theta)\right] = 1$ for the proof of Theorem 4.2.2. Let $Z_\theta = dP/dP_\theta$ be the ratio of the true underlying measure P with respect to the parametric measure P_θ corresponding to the constant parameter θ. Then,

$$\log Z_\theta = \sum_t \log \frac{g(X_t, X_{t-1}, \rho_t)}{g(X_t, X_{t-1}, \rho)},$$

where $g(X_t, X_{t-1}, \rho_t)$ is the density function for AFAR model with operator ρ_t corresponding to parameter θ_t and $g(X_t, X_{t-1}, \rho)$ is the density for a stationary FAR model with operator ρ corresponding to parameter θ. Similarly, on an interval I, we denote the probability measures by P_I and $P_{I,\theta}$ and we have

$$\log Z_{I,\theta} = \log \frac{dP_I}{dP_{I,\theta}} = \sum_{t \in I} \log \frac{g(X_t, X_{t-1}, \rho_t)}{g(X_t, X_{t-1}, \rho)}.$$

Then we obtain

$$\begin{aligned} E_\theta(Z_{I,\theta} \log Z_{I,\theta}) &= E(\log Z_{I,\theta}) \\ &= E\left[\sum_{t \in I} \log \frac{g(X_t, X_{t-1}, \rho_t)}{g(X_t, X_{t-1}, \rho)}\right] \\ &= E\left[\sum_{t \in I} E\big(\log \frac{g(X_t, X_{t-1}, \rho_t)}{g(X_t, X_{t-1}, \rho)}\Big|\mathscr{F}_{t-1}\big)\right] \\ &= E[\Delta_I(\theta)] \leq \Delta \end{aligned}$$

and the result follows, where $\mathscr{F}_{t-1}$ is the filtration or σ-field generated by all the past observations before time point t. □

Appendix D: Proof of Theorem 4.2.4

Proof. The first inequality follows from Corollary 4.2.3.

For the second inequality, we can derive the result by using inequality (4.10) and the property $x \geq \log x$ for $x > 0$. □

Bibliography

Afonso, A., Baxa, J. and Slavik, M. (2011). Fiscal development and financial stress: A threshold VAR analysis, *Technical report*, European Central Bank. Working paper series, No. 1319.

Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics* **17**: 579–600.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes, *Journal of Multivariate Analysis* **87**: 133–158.

Atanasova, C. (2003). Credit market imperfections and business cycle dynamics: A nonlinear approach, *Studies in Nonlinear Dynamics and Econometrics* **7**: 1–22.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**: 191 – 222.

Baillie, R. T. and Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach, *Journal of Economics Dynamics and Control* **33**: 1577–1592.

Balke, N. (2000). Credit and economic activity: Credit regimes and nonlinear propagation of shocks, *Review of Economics and Statistics* **82**: 344–349.

Bansal, R. and Zhou, H. (2002). Term structure of interest rates with regime shifts, *Journal of Finance* **57**(5): 1997–2043.

Baumeister, C., Durinck, E. and Peersman, G. (2008). Liquidity, inflation and asset prices in a time-varying framework for the euro area, *Technical report*, National Bank of Belgium.

Belomestny, D. and Spokoiny, V. (2007). Spatial aggregation of local likelihood estimates with applications to classification, *The Annals of Statistics* **35**: 2287–2311.

Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components, *Annals of Statistics* **37**: 1–34.

Besse, P. and Cardot, H. (1996). Approximation spline de la prévision d'um processus fonctionnel autorégressif d'order 1, *Canadian Journal of Statistics* **24**: 467–487.

Besse, P., Cardot, H. and Stephenson, D. (2000). Autoregressive forecasting of some climatic variations, *Scandinavian Journal of Statistics* **27**: 673–687.

Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes, *In: Roussas, G. (Ed.), Nonparametric Functional Estimation and Related Topics, NATO Science Series C*, Springer-Verlag, pp. 509–529.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, Springer-Verlag, New York.

Chen, J. and Gupta, A. (1997). Testing and locating variance changepoints with application to stock prices, *Journal of the American Statistical Association* **92**: 739–747.

Chen, L. (1996). Stochastic mean and stochastic volatility: A three factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives, *Financial Markets, Institutions and Instruments* **5**: 1–88.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models, *In: James J. Heckman and Edward E. Leamer (Ed.), Handbook of Econometrics*, Elsevier.

Chen, Y., Härdle, W. and Pigorsch, U. (2010). Localized realized volatility modelling, *Journal of American Statistical Association* **105**: 1376–1393.

Chen, Y. and Li, B. (2011). Forecasting yield curves in an adaptive framework, *Central European Journal of Economic Modeling and Econometrics* **3**(4): 237–259.

Chen, Y., Li, B. and Niu, L. (2013). A local vector autoregressive framework and its applications to multivariate time series monitoring and forecasting, *STATISTICS AND ITS INTERFACE* **6**(4): 499–509.

Chen, Y. and Niu, L. (2014). Adaptive dynamic nelson–siegel term structure model with applications, *Journal of Econometrics* **180**(1): 98–115.

Chen, Y. and Spokoiny, V. (2013). Modeling nonstationary and leptokurtic financial time series, *Econometrics Theory* . accepted.

Čížek, P., Härdle, W. and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models, *The Econometrics Journal* **12**: 248 – 271.

Cochrane, J. and Piazzesi, M. (2005). Bond risk premia, *American Economic Review* **95**: 138–160.

Cogley, T. and Sargent, T. J. (2001). Evolving Post-World War II U.S. inflation dynamics, *NBER Macroeconomics Annual* **16**: 331–373.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the Post-WWII U.S., *Review of Economic Dynamics* **8**: 262–302.

D'Agostino, A., Gambetti, L. and Giannone, D. (2013). Macroeconomic forecasting and structural change, *Journal of Applied Econometrics* **28**(1): 82–101.

Dai, Q., Singleton, K. and Yang, W. (2004). Predictability of bond risk premia and affine term structure models, *Working paper*, Stanford University.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models, *Biometrika* **82**(2): 339–350.

Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles, *Technical report*, Staff Report, Federal Reserve Bank of New York.

Diebold, F. X. (1986). Comment on: "modeling the persistence of conditional variance", *Econometric Reviews* **5**: 51–56.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching, *Journal of Econometrics* **105**: 131–159.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates, *Mathematical Finance* **6**: 379–406.

Egorov, A. V., Li, H. and Ng, D. (2011). A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates, *Journal of Econometrics* **162**: 55–70.

Eichler, M., Motta, G. and Von Sachs, R. (2011). Fitting dynamic factor models to non-stationary time series, *Journal of Econometrics* **163**(1): 51–70.

Engle, R. F., Ghysels, E. and Sohn, B. (2008). On the economic sources of stock market volatility. AFA 2008 New Orleans Meetings Paper.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica* **55**: 251–276.

Engle, R. F. and Rangel, J. G. (2008). The spline-garch model for low frequency volatility and its global macroeconomic causes, *Review of Financial Studies* **21**: 1187 – 1222.

Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association* **76**: 774 – 781.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates, *American Economic Review* **77**: 680–692.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.

Fernández-Rodríguez, F. (2006). Interest rate term structure modeling using free–knot splines, *Journal of Business* **79**: 3083–3099.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*, Springer-Verlag, New York.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation, *The Review of Economics and Statistics* **82**: 540 – 554.

Galí, J. and Gambetti, L. (2009). On the sources of the great moderation, *American Economic Journal: Macroeconomics* **1**(1): 26–57.

Geman, S. and Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves, *Annals of Statistics* **10**: 401–414.

Geweke, J. (1977). The dynamic factor analysis of economic time series models, pp. pp.365–383. In "Latent Variables in Socio-Economic Models", eds. D.J. Aigner and A.S. Goldberger. North-Holland, Amsterdam.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* **14**: 227–238.

Granger, C. W. J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* **11**: 399–421.

Granger, C. W. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**: 5–39.

Grenander, U. (1981). *Abstract Inference*, Wiley, New York.

Guidolin, M. and Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach, *Journal of Econometrics* **150**: 297–311.

Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive hilbertian processes, *Statistics & Probability Letters* **55**: 281–291.

Gürkaynak, R. S., Sack, B. and Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present, *Journal of Monetary Economics* **54**: 2291–2304.

Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992). A cointegration analysis of treasury bill yields, *The Review of Economics and Statistics* **74**: 116–126.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* **64**: 307–333.

Härdle, W. K., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer, New York.

Horváth, L., Husková, M. and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process, *Journal of Multivariate Analysis* **101**: 352–367.

Hosking, J. R. M. (1981). Fractional differencing, *Biometrika* **68**: 165–176.

Hull, J. and White, A. (1994). Numerical procedures for implementing term structure models ii: Two factor models, *Journal of Derivatives* **2**: 37–48.

Hwang, C. R. (1980). Gaussian measure of large balls in a hilbert space, *Proceedings of the American Mathematical Society* **78**: 107–110.

Jarrow, R., Ruppert, D. and Yu, Y. (2004). Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model, *Journal of the American Statistical Association* **99**: 57–66.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression, *Journal of Multivariate Analysis* **99**: 2508–2526.

Kokoszka, P. and Zhang, X. (2010). Improved estimation of the kernel of the functional autore- gressive process, *Technical report*, Utah State University.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change and the GARCH model, *Journal of Business & Economic Statistics* **8**: 225–234.

Li, F. and St-Amant, P. (2010). Financial stress, monetary policy, and economic activity, *Bank of Canada Review* **2010**: 9–18.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns, *Journal of Fixed Income* **1**: 54–61.

Liu, C. and Maheu, J. M. (2008). Are there structural breaks in realized volatility?, *Journal of Financial Econometrics* **6**: 326–360.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Mas, A. and Pumo, B. (2007). The ARHD process, *Journal of Statistical Planning and Inference* **137**: 538–553.

Mas, A. and Pumo, B. (2011). Linear processes for functional data, *In: F. Ferraty and Y. Romain (Ed.), The Oxford Handbook of Functional Data*, Springer-Verlag, pp. 47–71.

McCulloch, J. H. (1971). Measuring the term structure of interest rates, *Journal of Business* **44**: 19–31.

McCulloch, J. H. (1975). The tax adjusted yield curve, *Journal of Finance* **30**: 811–829.

Mikosch, T. and Stărică, C. (1998). Change of structure in financial time series, long range dependence and the garch model, *Technical report*, University of Groningen.

Mikosch, T. and Stărică, C. (2004a). Changes of structure in financial time series and the GARCH model, *REVSTAT Statistical Journal* **2**: 41–73.

Mikosch, T. and Stărică, C. (2004b). Non-stationarities in financial time series, the long range dependence and the IGARCH effects, *Review of Economics and Statistics* **86**: 378–390.

Mourid, T. and Bensmain, N. (2006). Sieves estimator of the operator of a functional autoregressive process, *Statistics & Probability Letters* **76**: 93–108.

Müller, H.-G. (2005). Functional modelling and classification of longitudinal data, *Scandinavian Journal of Statistics* **32**: 223–240.

Nakajima, J. and Ginkō, N. (2011). Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications, *Technical report*, Institute for Monetary and Economic Studies, Bank of Japan.

Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curve, *Journal of Business* **60**: 473–489.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors, *Biometrika* **95**(2): 365–379.

Peña, D. and Box, E. (1987). Identifying a simplifying structure in time series, *Journal of the American Statistical Association* **82**: 836 – 843.

Peña, D. and Poncela, P. (2004). Forecasting with non-stationary dynamic factor models, *Journal of Econometrics* **119**: 291 – 321.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies* **72**(3): 821–852.

Pumo, B. (1998). Prediction of continuous time processes by c[0,1]-valued autoregressive process, *Statistical Inference for Stochastic Processes* **3**: 297–309.

Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*, 2 edn, Springer, New York.

Schaefer, S. M. (1973). On measuring the term structure of interest rates, *discussion paper*, London Business School.

Schaefer, S. M. and Schwartz, E. (1984). A two factor model of the term structure: An approximate analytical solution, *Journal of Financial and Quantitative Analysis* **19**: 413–424.

Scharth, M. and Medeiros, M. C. (2009). Asymmetric effects and long memory in the volatility of dow jones stocks, *International Journal of Forecasting* **25**: 304–327.

Shea, G. S. (1985). Interest rate term structure estimation with exponential splines: A note, *Journal of Finance* **40**: 319–325.

So, M. K. P., Lam, K. and Li, W. K. (1998). A stochastic volatility model with markov switching, *Journal of Business & Economic Statistics* **16**: 244–253.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk, *The Annals of Statistics* **37**(3): 1405–1436.

Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns, *Review of economics and statistics* **87**(3): 503–522.

Svensson, L. (1995). Estimating forward interest rates with the extended Nelson & Siegel method, *Penning- & Valutapolitik* **3**: 13–26.

Tong, H. (1978). *On a threshold model*, In: Chen, C. (ed.) Pattern Recognition and Signal Processing. NATO ASI Series E: Applied Sc.(29). Sijthoff & Noordhoff, Netherlands, pp. 575-586.

Tong, H. (1983). *Threshold models in non-linear time series analysis. Lecture notes in statistics, No. 21*, Springer-Verlag, New York.

Tong, H. (1987). *Non-linear time series models of regularly sampled data: a review*, Vol. 2, VNU Science Press.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data, *Journal of the Royal Statistical Society B* **42**(3): 245–292.

Tsay, R. S. (1984). Order selection in nonstationary autoregressive models, *The Annals of Statistics* **12**(4): 1425–1433.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**: 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* **93**: 1188–1202.

Tsay, R. S. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models, *Journal of the American Statistical Association* **79**(385): 84–96.

Vasicek, O. A. and Fong, H. G. (1982). Term structure modeling using exponential splines, *Journal of Finance* **37**: 339–356.

Vetzal, K. R. (1994). A survey of stochastic continuous time models of the term structure of interest rates, *Insurance: Mathematics and Economics* **14**: 139–161.

Watanabe, T. and Omori, Y. (2004). A multi-move sampler for estimating non-gaussian time series models: Comments on Shephard & Pitt (1997), *Biometrika* **91**(1): 246–248.

Weron, R. (2007). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley.

Weron, R. and Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models, *International Journal of Forecasting* **24**: 744–763.